Use of Molecular and Bioinformatics Tools for Developing Methods of Epidemiological Identification of Trypanosomes

By Gaber Adwik

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

Jan 2016

College of science &Technology
School of Environment & Life Sciences
University of Salford

Contents

List of Figures

List of Tables

Acknowledgment

First of all I would like to thank my supervisor Professor Geoff Hide, as he collectively have provided me with a unique and complimentary mix of influences and assets. His support was an absolute help to me during my research. Professor Geoff always listens and his advice was always a solution for the issues I faced during my research.

I would also like to thank Dr Belgees Boufana for her advice and encouragement in the labs, and her ability to discuss and solve problems related to the devices in the labs. Also, I would like to thank Dr David Clough from University of Sheffield for reading my thesis for proof reading purposes

This thesis is dedicated to my mother and father, who works very hard to support me during my educational career and throughout my life.

Finally I would also like to thank my family and friends for their moral support during my research.

This research was financially sponsored by the Libyan Embassy in London, I would like to thank them for their financial support and solving my other problems. And I hope I can repay my country Libya for what it has giving me for the past years

Declaration Form

This is to certify that the thesis comprises only my original work towards the PhD. Due acknowledgement has been made in the text to all other material used. This work has not been submitted for any other degree or professional qualification

Gaber Adwik

……

Abbreviations

bp	Base pair
CATT	Card agglutination test for trypanosomiasis
CSF	Cerebrospinal fluid
DND*i*	*Drugs for Neglected Diseases Initiative*
ESAG	Expression site associated genes
FTA	Flinders technology Australia
HAT	Human African Trypanosomiasis
HCT	haematocrit centrifugation technique
ITS	Internal Transcribed subunit
LAMP	Loop-Mediated Isothermal Amplification
MDA	Multiple displacement amplification
MGE	Mobile Genetic Elements
NCBI	National Centre of Biotechnology Information
PCR	Polymerase chain reaction
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction fragment length polymorphism
RHS	retrotransposon hot spot
RIME	Ribosomal Insertion Mobile Element
SRA	serum resistance associated
TE	Transposable Element
TgsGP	*Trypanosoma brucei gambiense* specific glycoprotein
WHO	World Health Organisation

Abstract

Human African trypanosomiasis (HAT), also known as sleeping sickness, has been a major health problem for populations in Africa caused by the *Trypanosoma brucei spp* parasite. Although, the current number of the reported cases is on the decrease, more efforts are required to try to control or eliminate the disease. The recent advances in molecular techniques have contributed towards identifying taxonomic groups at all levels (species, subspecies, populations, strains and isolates). Commonly, field samples are collected and stored using Whatman FTA cards. Many molecular epidemiological tools are available for detection and strain typing in trypanosomes. These tools include nested ITS-PCR, which is based on size variation of the ITS genes and MGE-PCR, which is based on variations in position of mobile genetic elements (MGEs). Although commonly used, these tools have not been fully validated. For example, the ITS-PCR has not been used or validated against blood samples obtained from sleeping sickness patients in Angola. Furthermore, the MGE-PCR system has not been evaluated for use directly from FTA cards.

The aim of this thesis is to develop improved molecular tools to assist diagnostic and epidemiological studies. In order to improve the molecular diagnostic use of Whatman FTA cards, an extraction method based on Chelex was investigated. Using Chelex to extract *T. brucei* DNA from FTA cards, followed by a nested ITS-PCR detection system, allowed parasite DNA detection to 1ng/µl. To evaluate this tool on field samples, ITS-PCR amplification was carried on DNA eluted by Chelex extraction from 36 FTA cards spotted with blood from Angolan patients tested positive for trypanosomiasis by the card agglutination test (CATT). Twenty four of these samples were successfully PCR amplified using mammalian tubulin primers. Of these 24 samples, 11 (= 45.8%) were confirmed as trypanosome positive utilising a specific

ITS-PCR based approach. As such, this indicates that further work is necessary to improve the PCR-based reliability of diagnosis.

To this end, an MGE-PCR approach was used to attempt parasite strain identification.. Although the MGE-PCR was found to be more sensitive than ITS-PCR in amplification of DNA from FTA cards, the resulting sequence data was not able to confirm that the amplicons were of trypanosome origin and hence further analysis, or approaches, are required.

With a view to developing new diagnostic tools a bioinformatic analysis of mobile elements inserted in *RHS*/pseudogenes in the *T.brucei* genome was carried out. The aims of this were to locate variable regions of these genes that could be used as detailed markers for trypanosome strain identification. Sequences of the *RHS* genes were retrieved from the *T.brucei brucei* and *T. brucei gambiense* genomes to investigate positional diversity of MGEs within this family of genes. Differences were found in the presence/absence of RIME elements in one *RHS* gene between the two subspecies. More detailed investigation of all *RHS* gene classes in *T. b. brucei* showed six classes of *RHS* gene types and within each class, individual sequences showed evidence of insertion by MGEs. In some specific instances, evidence of pre-insertion, insertion and subsequent removal of MGEs was seen. This enabled a temporal evolutionary sequence of events to be interpreted. As such, the *RHS* genes offer the opportunity to develop specific molecular epidemiological tools for investigating the evolution of MGEs in field samples and to carry out temporally informed epidemiological tracking of isolates.

Chapter 1 **Introduction**

Human African Trypanosomiasis (HAT) which is also known as sleeping sickness is a disease was associated with sleeping inappropriately. This is due to neurological disturbances associated with the second stage of the disease, when parasites have become manifest in the central nervous system (Barrett, 1999; Kennedy, 2005). The African trypanosomiases comprise a group of economically important animal diseases and medically important zoonotic diseases that affect much of sub Saharan Africa. With economic losses in cattle production alone being in the range of US$ 1.0 - 1.2 billion (Daffala *et al.,* 2013). HAT is considered as a serious public health problem that is caused by trypanosomes, protozoan parasites, transmitted by tsetse flies (*Glossina*). However, it is not considered as important on the world stage as diseases such as malaria and AIDS, but it is an important disease and responsible for an extensive degree of suffering and mortality in countries where it is endemic (Hide, 1999).

In addition, it occurs in 36 African countries between latitudes 14˚ North and 29° south. Following the geographical distribution of the tsetse fly, more than 50 million people living in some 200 foci in this area are at risk of acquiring the infection (Kuzoe, 1993).

The World Health Organisation (WHO) estimates suggest that 500,000 new cases of sleeping sickness occur each year with about 60 million people considered at risk, variations in the annual number of reported cases must be interpreted with caution due to the impact of the civil war on case-finding (TDR, 2004). Furthermore, according to TDR, (2002) less than 4 million people are under surveillance and only about 40,000 are diagnosed and treated.

However, thanks to an ambitious campaign led by WHO, many nongovernmental

organizations and a public private partnership with Sanofi- Aventis and Bayer donates and distributes the necessary drugs to WHO for use in affected countries (Simarro *et al.*, 2011a). According to WHO (2012) in 2012 the number of reported cases was fewer than 8000, thus WHO have designed a road map to eliminate sleeping sickness as a public health problem by 2020.

1.1 History of human sleeping sickness (HAT)

The first report of human sleeping sickness is ascribed to Arabic historians describing the death of King Diata II, Sultan of Mali in 1373 (De Raadt, 1999), whereas, nagana of cattle has been known since 15th century (Nash, 1969). At the beginning of 1900, the importance of trypanosomes as pathogens was recognized by Timothy Richards Lewis and Griffith Evans in India and by Sir David Bruce in East Africa all of whom gave their names to species of trypanosomes (Goodwin, 1985). However, the first time that a trypanosome infection was reported in humans was by Forde in the Gambia in 1901 (Forde, 1902) he managed to find the parasites in the blood of a European suffering from recurrent fevers. Furthermore, between 1900 and 1915 Portuguese, British, French, German and Belgian scientific teams were sent out to Africa. Their aim was to study the cause, the factors which help the spreading of the disease, and the severity of the problem in the respective colonies. Since then sleeping sickness has become part of the domains of biological and medical science (WHO, 2005).

On the other hand, between 1895 and 1910 there were serious epidemics in several regions, mainly in the Congo basin, especially northern Angola and around the Lake Victoria basin for this epidemic period the total number of deaths due to trypanosomiasis in the Congo basin was estimated at 500,000 and in the Lake Victoria region at around 250,000 (TDR, 2004).

Moreover, by the end of 19th century, the colonial authorities in Africa had become more seriously concerned about the trypanosomiasis problems (livestock and human disease). Between 1939 and 1960 mobile teams in French colonies of west and central Africa had examined 73,000,000 individuals and detected 250,000 cases (Pépin and Méda, 2001). In the 1960s, tsetse and trypanosomiasis research and control services were well established throughout most of the endemic regions of Africa. The number of cases of human sleeping sickness decreased to an estimated level of no more than 15,000 and 25,000 per year for the whole continent (TDR, 2004).

1.2 The pathogenic parasite- *Trypanosoma brucei*

The trypanosome is a spindle-shaped cell (20 to 30 by 1.5 to 3.5 nm) with a single flagellum. The flagellum is structured as the posterior end, runs along the cell membrane, to which it is attached by an undulating membrane, and extends beyond the anterior part of the cell. The base of the flagellum is associated with the kinetoplast, a large particle containing the DNA of the single mitochondrion (Chappuis *et al.*, 2005).

Although there are many species of trypanosome, several species cause important veterinary diseases, and only two cause significant human diseases. In sub-Saharan Africa, *Trypanosoma brucei* causes sleeping sickness, and in America, *Trypanosoma cruzi* causes Chagas disease, transmission of both species can also be via blood transfusion, contaminated needles, or the congenital route.
Rarely, transmission of *T cruzi* by breastfeeding or through contaminated food has been reported (Barrett *et al.*, 2003). On the other hand, cattle infections caused by *T. congolense* and *T. vivax* are by far the most serious, both for frequency and for economic influence (Flneile *et al*., 1983). Moreover, *T. brucei* is taxonomically divided into three subspecies; largely based on their geographical origin, infectivity to

humans and severity of disease *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense*.

T. b. brucei is distributed across sub-Saharan Africa, and is restricted to non-human vertebrates, in which it can cause nagana, a chronic wasting disease (Hoare, 1972). *T. b. gambiense* is restricted to West and Central Africa, where it causes a chronic form of sleeping sickness in humans the Gambian form of sleeping sickness caused by. *T. b. gambiense*, was traditionally viewed as primarily a human infection, but it has become clear that a broad range of wild and domestic animal reservoirs also harbor the parasite (Mehlitz *et al*., 1982; Cordon-Obra *et al*., 2009). *T. b. rhodesiense* is a human infective subspecies and is found in eastern and southern Africa and causes an acute form of sleeping sickness. *T. b. rhodesiense* is a zoonotic disease for which non-human vertebrates are the primary reservoir (Hoare, 1972). Since *T. b. gambiense* is less virulent than *T. b. rhodesiense* there is a widespread belief that *T. b. gambiense* is evolutionarily older than *T. b. rhodesiense* and *T. b. brucei* based on the hypothesis that parasites generally evolve towards becoming more benign as they adapt to their host, an assumption that is not necessarily supported by evidence (Bull, 1994; Ebert *et al*., 1996).

Besides this, another hypothesis is that *T. b. rhodesiense* may evolve through frequent genetic exchange with sympatric *T. b. brucei*, leading to a mosaic of different *T. b. rhodesiense* genotypes distributed throughout endemic regions of eastern Africa (Komba et *al*., 1997; Gibson *et al*., 2002). Laboratory studies have demonstrated that *T. brucei* is capable of sexual reproduction (Jenni *et al*., 1986) and that crosses between *T. b. brucei* and *T. b. rhodesiense* can produce viable progeny (Gibson, 1985). Although, *T. b. gambiense* and *T. b. rhodesiense* shares 99% of the genome sequence identity with *T. b. brucei* (Jackson *et al*., 2010), each has evolved distinct mechanisms to escape lysis by human serum; these are the

causative agents of human African trypanosomiasis (Simarro *et al*., 2011a). In vitro tests, in which trypanosomes are incubated with human blood or serum (e.g. blood-incubation- infectivity test (Rickman and Robson, 1970) or human-serum-resistance test (Jenni and Brun., 1980) are based on the trypanosome lytic factors (TLFs) present in human blood, which can lyse *T. b.brucei* but not *T.b.rhodesiense* (due to the presence of human-serum-resistance-associated (*SRA*)) or *T.brucei.gambiense* (Raper *et al*., 2001). Surprisingly, African trypanosomes have more than one mechanism of human-serum resistance, because *SRA* was not detectable in *T. b. gambiense* by analysis of mRNA or DNA (Wellburn *et al*., 2001) Attempts to discover the corresponding mechanism in *T.b.gambiense* have not yet fully succeeded, although similar approaches to that used to discover the SRA gene in *T.b.rhodesiense* have been tried (Berberof *et al*., 2001). However, Berberof *et al.* (2001) have reported a gene which is absolutely specific to *T. b. gambiense*, termed as *T. gambiense-specific glycoprotein* (*TgsGP*). In contrast to *SRA*, *TgsGP* cannot confer normal human serum, progressive deletion of the *TgsGP* sub-telomeric region revealed that *TgsGP* is necessary for resistance to normal human serum (Uzureau *et al*., 2013). Genetic analysis indicated that *T. b. brucei* is closely related to *T. b. rhodesiense*. The hypothesis of a common ancestor for both *T. b. rhodesiense* and *T. b. brucei* was mentioned (Gibson *et al*., 1999).

T. b. gambiense (chronic) is responsible for 95% of the current cases of HAT (WHO, 2012). Rhodesiense HAT has epidemic potential in humans, as it has been responsible for large outbreaks in the past (Köerner *et al*., 1995; Hide, 1999). As between 1900 and 1915 *T. b. rhodesiense* recorded the largest epidemic killed more than 250 000 people over 15 years (Langlands, 1999). Due to different factors, the geographical gap between the two forms has been reducing in Northwest Uganda and on the border between the United Republic of Tanzania and

the Democratic Republic of Congo, and in the future, an overlap of the two forms could occur (Piccozzi *et al*., 2005). The three subspecies share all the morphological features, which make the diagnosis and identification harder. Also these taxonomic criteria have some gaps, for instance it is clear that both *T. b. gambiense* and T*. b. rhodesiense* have animal reservoir hosts, thus by implication an isolate from animal is not necessarily non-human infective (Hide, 1999).

T.b. gambiense is fatal which may lead to death if left untreated. In addition, it is found as two stages; first stage and second stage. First stage starts from the bite of testse fly which characterized chronic and intermittent fever, headache, pruritus, lymphadenopathy and, to a lesser extent, hepato-splenomegaly. Whereas, in the second stage is characterized by sleeping disturbances, neuro-psychiatric disorders and the rare presence of fever. Neurological symptoms in second stage is described by tremor, fasciculation, general motor weakness, paralysis of an extremity, hemiparesis, akinesia and abnormal movements such as dyskinesia or choreoathetosi (Kennedy, 2006).

1.3 Life cycle of *T. brucei*

Trypanosome life cycles can be considered relatively complex, and can be divided into two events in the tsetse fly and inside the mammalian (figure 1-1) (Vickerman, 1985; Dyer *et al*., 2013). Firstly, the life cycle of the trypanosome inside the mammalian host; the life-cycle begins when the tsetse fly injects the metacyclic forms into the mammalian host, and then starts the adaption phase adapted for life in the bloodstream inside a tsetse. The metacyclic form is morphologically characterized including differentiating and proliferating into; long, slender bloodstream forms known as trypomastigote forms (infective form) (Vickerman, 1985). During this form diagnoses can be carried in accordance with the presence of

the variant surface glycoprotein (VSG) coat that will protect the parasite in order to survive in the host. (Despommier *et al.*, 2005; MacGregor *et al*., 2010). Then the blood form can enter into different body fluids, such as lymph and cerebrospinal fluid, and can also enter the placenta (WHO, 2013). From the fluids the parasite will migrate to the organs particularly the brain and central nervous system (CNS). *T. b. rhodesiense* establishes itself within the CNS within a few weeks (Odiit *et al*., 1997) while *T. b. gambiense* takes an average of 300–500 days to manifest in the CNS (Checchi *et al*., 2008a).

On the other hand, events occurring inside the vector begins when the tsetse fly takes a blood meal and the parasites are in bloodstream trypomastigote forms, then they migrate to the midgut. Next, once they arrive to the midgut, trypomastigote forms start to differentiate, via the oesophagus, proboscis, and hypopharynx they migrate to the salivary gland, where they are able to multiply and some of them can transform into infectious metacyclic forms (Sharma *et al*., 2009). During this migration from the midgut to the salivary glands, the parasite population size experiences a pronounced reduction (Oberle *et al*., 2010). The whole cycle in the vector takes 18–35 days, and once infected, a tsetse fly remains so for the rest of its lifespan (Vickerman *et al*., 2003).

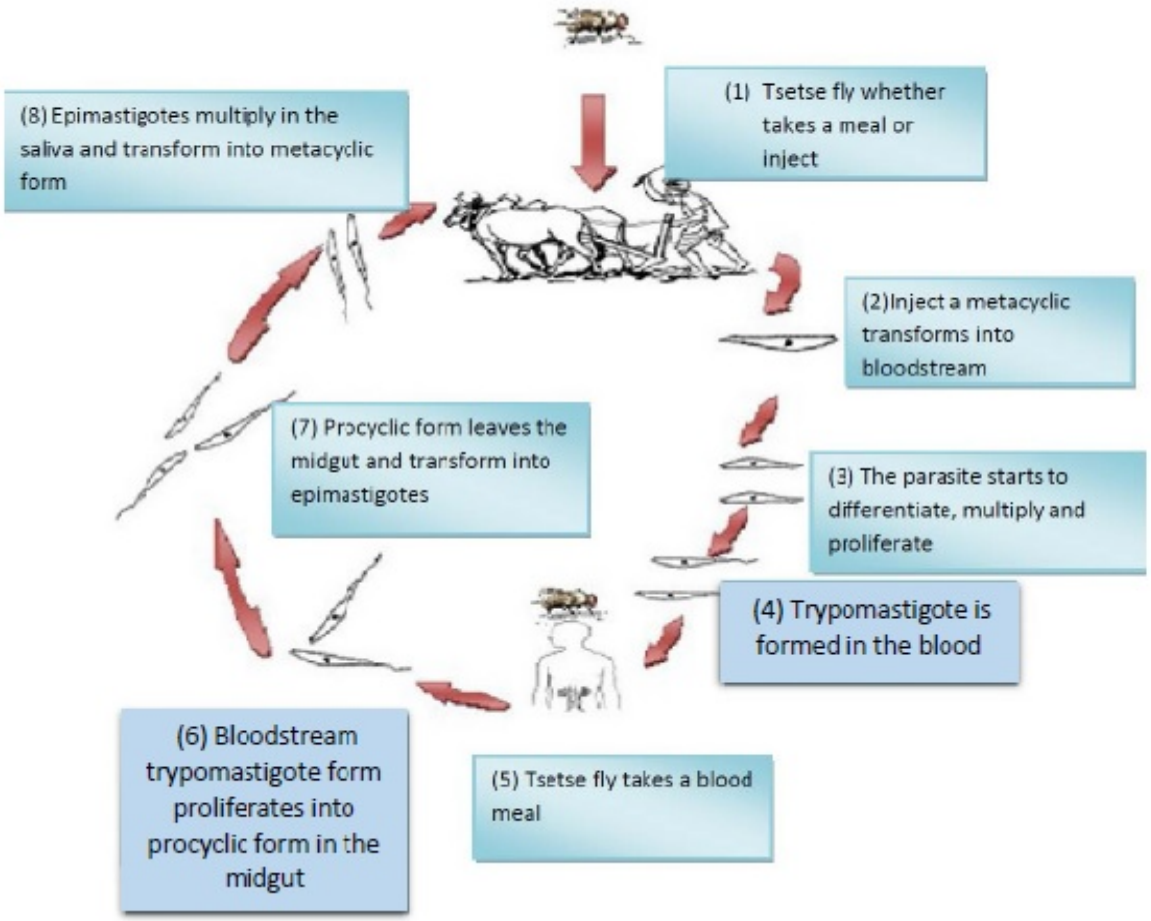


Figure 1-1. The life cycle of *T. brucei* in the vector and in the host; numbers represent the number of stages in order

1.4 Transmission cycle

According to (WHO, 2013) HAT has two forms of transmission according to the form of the disease whether it is chronic that caused by *T. b. gambiense,* or it is acute form, caused by *T. b. rhodesiene.* Gambiense HAT is an infectious disease that involves a minor role for animal reservoirs, while Rhodesiense HAT is a zoonotic disease, affecting mainly animals; humans are only accidental hosts (Franco *et al*., 2014a). Although, the tsetse fly has a considerably low capability as vector for gambiense HAT, also the amount of parasitemias in human is usually low; the long duration of human infection with an extended few symptoms period is considered sufficient to maintain a human–fly–human

transmission cycle (Robays *et al*., 2004). However, rarely animals can play role in transmission of Gambiense HAT. As Njiokou *et al*., 2010 described that *T.b.gambiense* was found in domestic and wild animals, shows the transmission cycle of the *T. b. gambiense*. Some studies show that experimental infections of animals with *T. b. gambiense* are limited in time, always lasting less than 1 year (Schütt *et al*., 1981). Therefore, more data are needed to clarify the role played by the animal reservoir in maintaining gambiense transmission when the prevalence of human infection is low and elimination is envisaged.

On the other hand, domestic and wild animals play a major role in *T. b. rhodesiense* sleeping sickness transmission cycle. Intensified man to man transmission occurs during epidemic. The strong zoonotic character of the *T. b. rhodesiense* form of the disease considerably complicates surveillance and control issues, requiring action on the fly or on the animals hosting the parasite.

1.5 The vector- Tsetse fly

Tsetse fly *(Glossina* spp.) is the only known biological vector of human trypanosomiasis. It can be described as brown in colour with an extended snout, the fly ranges from one-quarter to two-thirds inches (0.6 to 1.7 cm) long, the smaller varieties being slightly larger than a housefly. Of the 30 known living tsetse species, only eleven are of positive or possible epidemiological importance as trypanosome vector (Knight, 1971). Tsetse flies are subdivided into three subgenera based on morphological differences in the structure of the genitalia: *Morsitans* (*Glossina*), *Palpalis* (*Nemorhina*), and *Fusca* (*Austenina)* groups (Despommier *et al.,* 2005). Tsetse flies are geographically restricted to sub-Saharan Africa, between the latitudes of 14°N and 29°S, south of the Sahel desert, and north of the Namibian and Kalahari deserts (Figure 1-2) (Fèvre *et al.*, 2006). Some pockets of *G. morsitans* and *G.fuscipes*

have been reported in south western Saudi Arabia (Elsen *et al*., 1991). All the species of tsetse flies are potentially cyclical vectors of trypanosomes, but in nature the infection is carried almost exclusively by *G. fuscipes*, *G. palpalis*, and *G. morsitans* (Franco *et al*., 2014a).

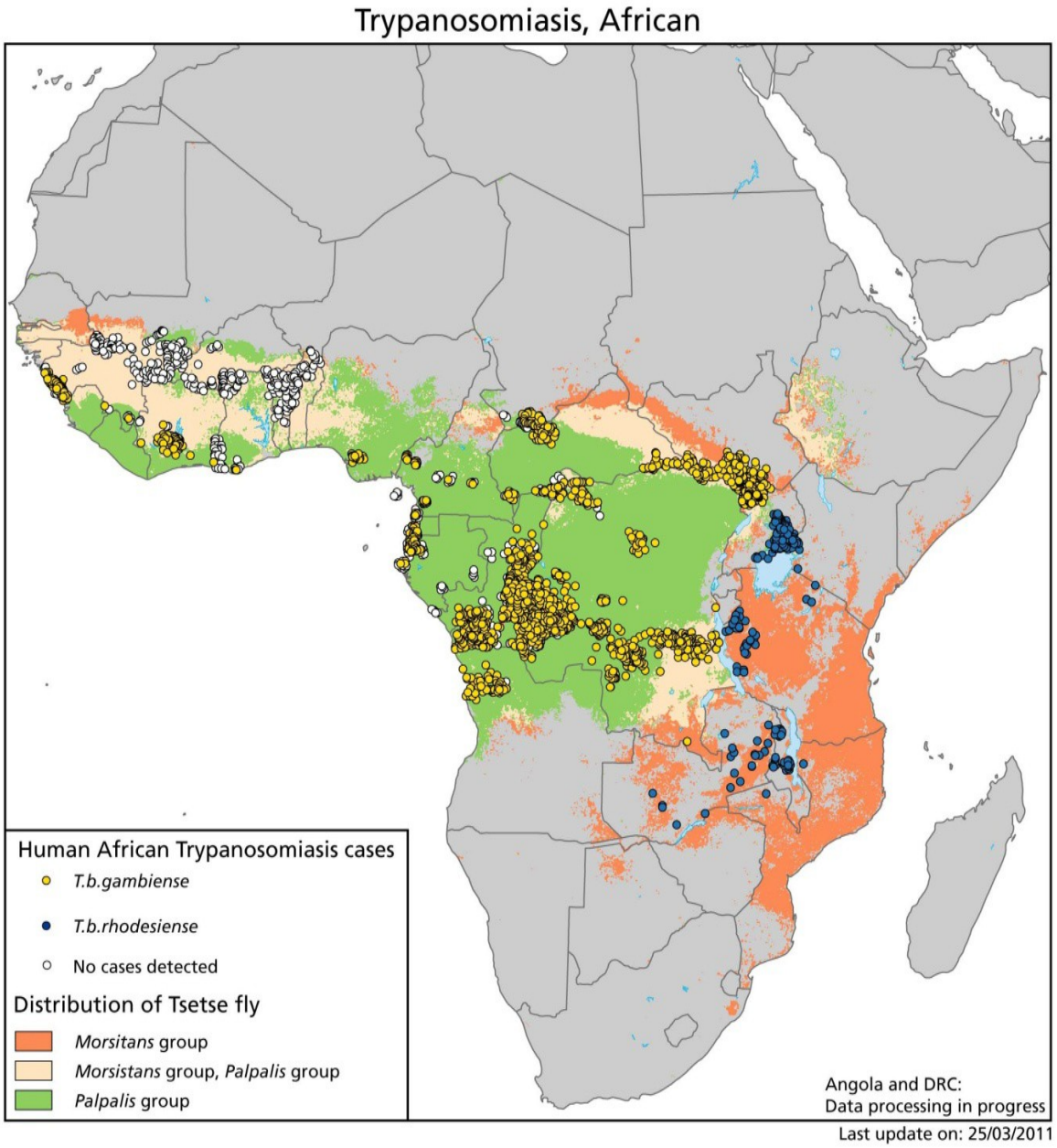


Figure 1-2. The distribution of *T. b. gambiense* (highlighted in yellow) and *T. b. rhodesiense* (highlighted in blue) and subgenera of tsetse fly in Africa

Source: Atlas of Human Infectious Diseases (2012)

Recently the full genome of *Glossina morsitans* has been fully sequenced (Attardo *et al*., 2014). This breakthrough will certainly help in controlling the disease, by

understanding the pathways of transmission particularly those which occur in the tsetse fly saliva. As the authors have reported a family of genes, the tsal genes, which are particularly active in the salivary glands of the tsetse fly. This allows the tsetse fly to counteract the responses from the host to stop bloodfeeding. This finding and several others could be useful in developing new tools that could reduce or even eradicate tsetse flies.

1.6 Epidemiology African sleeping sickness

The first reported isolation of *T. b. rhodesiense* from a sleeping sickness patient was in 1908 (Hide, 1999). On the other hand, *T. b. gambiense* HAT at the turn of the twentieth century (1896– 1906), is estimated to have killed 800 000 people in equatorial Africa. The disease (HAT) resurged in the 1920s, threatening both West and Central Africa; the *T. b. rhodesiense* outbreak lasted for a decade in some localized areas and for much longer in larger countries. The epidemics are raging in the worn- torn areas of southern Sudan, northern Uganda, the Democratic Republic of Congo, and Angola (Hide, 1999). (Figure 1-3) shows the distribution of HAT across Africa reported in 2013.

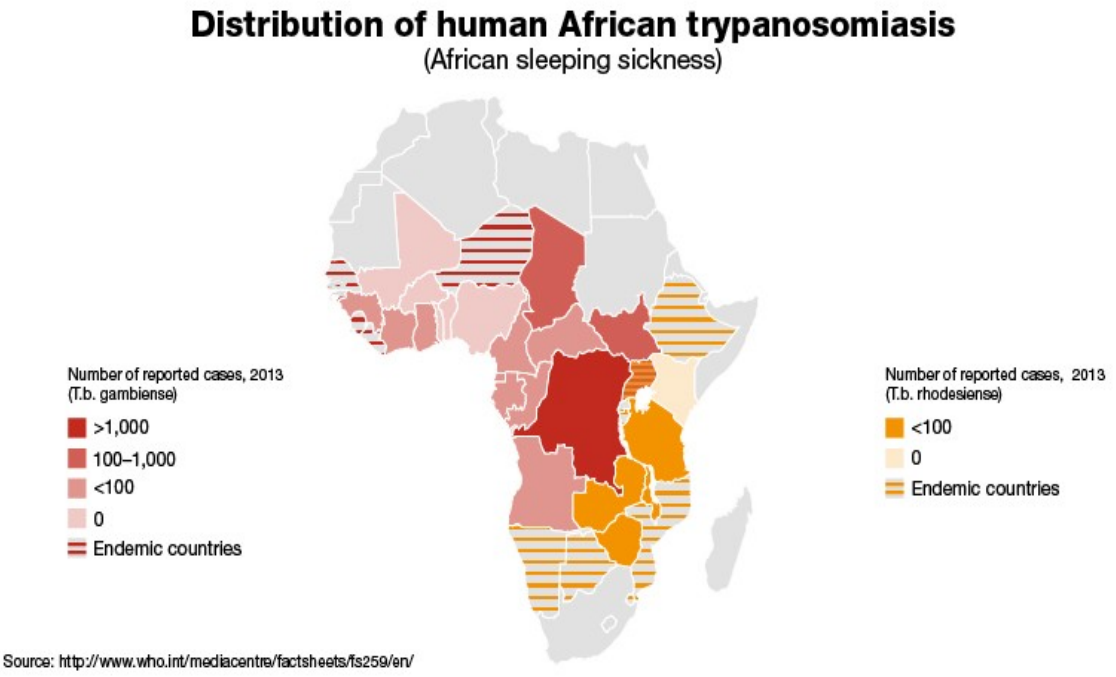


Figure 1-3. The number of HAT (*T.b.gambiense* and *T.b.rhodesiense*) cases and the geographical distribution in 2013.

Source: http://www.who.int/mediacentre/factsheets/fs259/en/ (Accessed on 20/12/14).

According to figure (1-3), out of the known 36 endemic countries, over 90% of the cases were reported from Angola, Democratic Republic of Congo, southern Sudan and Uganda. Thus, *T. b. gambiense* trypanosomiasis is considered one of the major public health problems in Central Africa, especially in the Democratic Republic of Congo (DRC), Angola and southern Sudan, where ongoing civil wars have negative effect on the control efforts to such an extent that national statistics give only a very incomplete view of the problem. Better information is available in DRC, the total number of people at risk is estimated by the national control programme to be 12,500,000. The WHO Expert Committee on HAT control and surveillance in 1995 estimated that the true prevalence of gambiense HAT was 300 000 cases. The number of cases reported annually increased noticeably from about 10,000 in 1980

to more than 27,000 in 1998 in DRC. About 2% of DRC's population carries the disease, with prevalence in some communities reaching 70% (Barrett, 1999).

Moreover, Angola is considered as the second highest incidence of HAT; respectively 8275 and 6610 new cases were reported in 1997 and 1998 by the national control programme. The disease is endemic in the north-west provinces, the prevalence rates reported vary between 1.3% and 9.7% (Barrett, 1999).

From 1997-2006 24 countries out of the 36 endemic countries showed greatest burden of the disease due to *T.b.gambiense* (WHO, 2006). In 2006 11 out of these 24 countries have reported no cases (Wellburn and Maudlin, 2012). These data, especially for *T.b. gambiense*, illustrate an encouraging trend for countries where concerted efforts have been mounted (mainly Angola, DRC, and Sudan) to control HAT and the decrease in incidence in recent years is due in large part to enormous efforts involving active case detection (Fèvre *et al*., 2008). (Figure1-4) illustrate the number of HAT cases reported from 1997 to 2013.

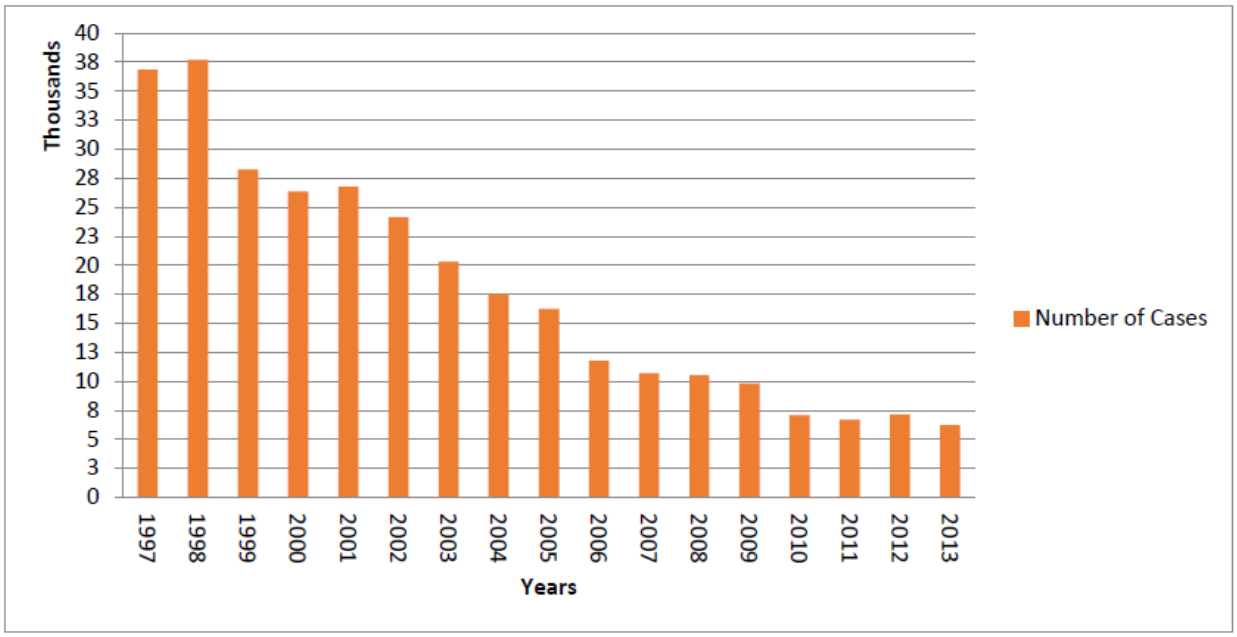


Figure 1-4. The number of cases of HAT reported from 1997-2013

(Simarro *et al.*, 2011; WHO, 2013)

Furthermore, Uganda is the only country where both sub-species are found as they are currently less than 100 km apart *T. b. gambiense* in the north –west and *T. b. rhodesiense* in the south-east. Between 1000 and 1200 *T .b. gambiense* cases were reported annually in 1995-2000. Whereas, *T. b. rhodesiense* cases it is less than *T .b. gambiense*, only 271 cases were reported in 1997. In addition, a major epidemic of Rhodesian HAT devastated south-east Uganda from the mid-1970s, with a cumulative total of about 40,000 new cases for 15 years (Pépin and Méda, 2001). Picozzi *et al.,* (2005) have screened 231 blood samples were collected from the central Uganda (2001-2005) and 91 from northeast of Uganda (2003). The screening was carried based on the detection of *SRA and TgsGP* genes. The authors have reported that all the samples from central Uganda are classified as *T.b.rhodesiense*, where as all the samples from northeast Uganda are classified as *T.b.gambiense*

Moreover, in Sudan where reliable statistics are not available, the foci of *T. b. gambiense* HAT are located in the southern part of the Equatorial region, west of the

Nile, within 100 kilometres of the borders with Central African Republic (CAR), the DRC and Uganda. Extrapolations suggest that there must be at least a few thousand cases per year. In other countries of Central Africa, HAT is more or less an emerging public health problem (Franco *et al*., 2014a).

The increase of gambiense HAT cases in the first third of the 20th century, which reached, in some cases, epidemic levels, generated a response from colonial governments that were initially based on the application of vector control measures, and later on the deployment of mobile teams that visited the villages in the at-risk areas to conduct systematic screening of the entire population and to treat the detected cases (Ekwanzala *et al*., 1996; Moore and Richer, 2001).

The trend in cases reported of gambiense HAT showed a decrease in the reported cases of the disease of 76% since 2000, with a number of around 7,000 cases reported in the last 3 years (WHO, 2013). Between 2000 and 2009, the annual case detection rate decreased by 63% at the same time as geographical and population coverage by active screening increased (Simarro *et al*., 2011a). This promising result continued in 2010 and 2011. In 2011, the number of new cases reported to WHO was 6631 (Simarro *et al*., 2013). The Democratic Republic of the Congo currently bears most of the disease burden. Nevertheless, HAT mainly affects remote rural communities where the health infrastructure is basic; furthermore, there are other endemic areas where accessibility is complicated because of security problems or topography constraints. Therefore, a certain number of cases are not recognized and diagnosed (Simarro *et al*., 2011a) despite the fact that epidemiological knowledge about the disease has improved considerably in the past decade there is still a gap between the number of cases declared and the number of actual cases.

Elsewhere in East and Southern Africa, the incidence of *T. b. rhodesiense* trypanosomiasis remains low. In West Africa, the disease has regressed or disappeared from several countries as ecological changes reduced the intensity of man-fly contact (Pépin and Méda, 2001). Reporting from most countries endemic *T . b. rhodesiense* trypanosomiasis is fairly irregular, as the disease is not regarded as priority. For instance Tanzania report between 200 and 400 cases per year, and there are several foci throughout the country. Additionally, lower figures are reported in Zambia and in Kenya, where the disease is seen essentially near the Ugandan border, in contrast fewer than 20 cases per year are reported in Rwanda, Mozambique, Malawi, Botswana, Zimbabwe (Pépin and Méda. 2001).

Furthermore, trypanosomiasis is not found equally in males and females, in some foci the incidence and prevalence is higher in males if the tsetse population is denser in areas of typically male activities (Pépin and Méda, 2001). Females often participate more regularly in case-finding sessions, which can lead to a higher number of cases (WHO, 2001). Which may causes a misleading if the estimation was only by the number of cases (Pépin and Méda, 2001).

On the other hand, children under the age of 10 years usually have a three to five folds lower incidence than adults, and the reason is the differential exposure, as tsetse populations are often more abundant in fishing and farming sites than in the villages. Furthermore, individuals over 50 to 60 years of age are less risk to trypanosomiasis, than in middle-aged adults. Occupation is also related to exposure and to incidence, for example in Côte d'Ivoire HAT is more frequent in coffee and cocoa plantation workers or in people who fetch water than in other inhabitants. Nearly 80% of cases occur in people who not only work but live in small plantation settlements; a case control study showed that such people were five times more likely to develop trypanosomiasis, than their counterpart who lived in villages (Pépin and Méda, 2001).

1.6.1 Sleeping sickness in Angola

The total population of Angola is about 24 million (World Bank, 2014) Since gaining independence in 1975, Angola has seen a cruel civil war, which over nearly 30 years, has devastated the country, destroyed the infrastructure and left a severely traumatized population (Abel *et al*., 2004). Daily conditions of life throughout the country and specifically Luanda (population approximately 6 million) mirror the collapse of administrative infrastructure as well as many social institutions. Attempts were made by the United Nations (UN) to promote a process of peace and reconciliation failed (United Nations Department of Public Information 1995), and only

after the death of the UNITA opposition leader Jonas Savimbi in February 2002, has raised a new glimpse of hope. The peace agreement signed on 4 April 2002 between the Angolan government and UNITA, which have initiated a new era of stability and progress for the country (Abel *et al*., 2004). Since the adoption of a new constitution in 2010, the politics of Angola takes place in a framework of a presidential republic, whereby the President of Angola is both head of state and head of government, and of a multi-party system. Executive power is exercised by the government. Legislative power is vested in the President, the government and parliament. The health system has witnessed signs of improvements (constitution net [online]).

Moreover, HAT was first considered as a health problem in Angola in 1871. The first reported case of HAT in Angola was in the

Quicama municipality province (Janssens, 1996). However, it took several decades for the fight against the disease to be systematically organized. By 1926, the Portuguese colonial government had created the first specialized unit for HAT control. In 1949, the first mobile team was formed to screen the population in the affected areas. Less than 10 years later, nearly the whole population in the endemic zones was under constant surveillance. At the same time, antitrypanosomal treatment became available (Abel *et al*., 2004). In 1974, a year before the independence, only three cases were reported the entire country (Stanghellini and Josenando, 2001). Subsequently, due to the civil war the number of population screening fell dramatically. The level of professional performance of health workers declined constantly, as the activities of nursing schools and regular training programmes for laboratory technicians gradually came to a standstill. As a result, HAT returned to Angola in a devastating dimension, as the number of cases have risen from 3 in 1974 (Abel *et al*., 2004) to 1274 in 1994 (WHO, 1998). However, a project was created in 1995 and started its activities in 1996 under the name ANGOTRIP, an acronym

combining the two words Angola and trypanosomiasis. As part of Caritas de Angola, ANGOTRIP could make use of the widespread and sustainable infrastructure of the Catholic Church, which was of great advantage even during highly volatile periods, when the UN and international NGOs were no longer able to reach the local population. However, it did not have a major impact on reducing the reported number of cases, as the number of cases has increased dramatically to 6610 in 1998 (Stanghellini and Josenando, 2001).

Since 2001 in Angola, 80% of patients have been found to be in the meningoencephalitic stage of the disease (Truc *et al*., 2012). Moreover, in 2001, during a resurgence of the infection, WHO and partners established a public–private partnership to bring the disease under control. This was effective within a short time, in terms of reducing the number of reported cases of HAT (Figure 1-4). Angola was not an exception as the number of cases fell to 4577 (WHO, 2013). Also, in 2004, therefore, the World Health Assembly passed a resolution (WHA57.2) to enhance control in order to move towards elimination. In 2011, in view of the 70% decrease in the number of cases reported to WHO in the previous decade**.** The resolution could be considered as a huge relief for Angolans as the number of reported cases kept falling**.** Finally, The Atlas was initiated in 2007 as a collaborative project between WHO and the Food and Agriculture Organization of the United Nations (FAO) in the framework of the programme against African trypanosomiasis. The Atlas is an excellent basis for monitoring changes in distribution and epidemiology and for using spatial analytical methods to estimate the size and location of populations at risk (WHO, 2013). Again Angola has majorly benefited from this project as the number of cases has decreased since the number of reported cases has fallen noticeably to 517 cases in 2008 (WHO, 2013). In the period from 2009–2013, 82% of gambiense HAT cases were reported from the

Democratic Republic of the Congo, 5.5% from the Central African Republic, 4% from Chad, 3.5% from South Sudan, 2% from Angola, and the remaining 3% from another five countries (Cameroon, Congo, Cote d'Ivoire, Equatorial Guinea, Gabon (Franco *et al*., 2014a). The current trends of the number of cases has shown promising signs of controlling the disease in Angola since fewer than 100 cases have been reported per year in Angola (Table 1-1). (WHO, 2015)

Country	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Angola	2280	1727	1105	648	517	247	211	154	70	69

Table 1-1. The number of cases reported of *T. b. gambiense* sleeping sickness between 2004 and 2013 in Angola (Simarro *et al.*, 2008; WHO, 2013).

Moreover, Angola (northwest region) is considered as high intensity transmission country along with Chad (southern region); the Central African Republic (Northwestern and Southeastern regions), Equatorial Guinea (Southwestern region), Gabon (Northwestern region), Congo (central and Southern regions), the Democratic Republic of the Congo (in Northern, Northwestern, central, and Eastern parts), South Sudan (Southwestern region), and Uganda (Northwestern region) (Simmaro *et al*., 2013). There are three subgenus of tsetse fly responsible for transmitting HAT which are; *G. palpalis*, *G. fuscipes* and *G. morsitans* (Figure 1-5).

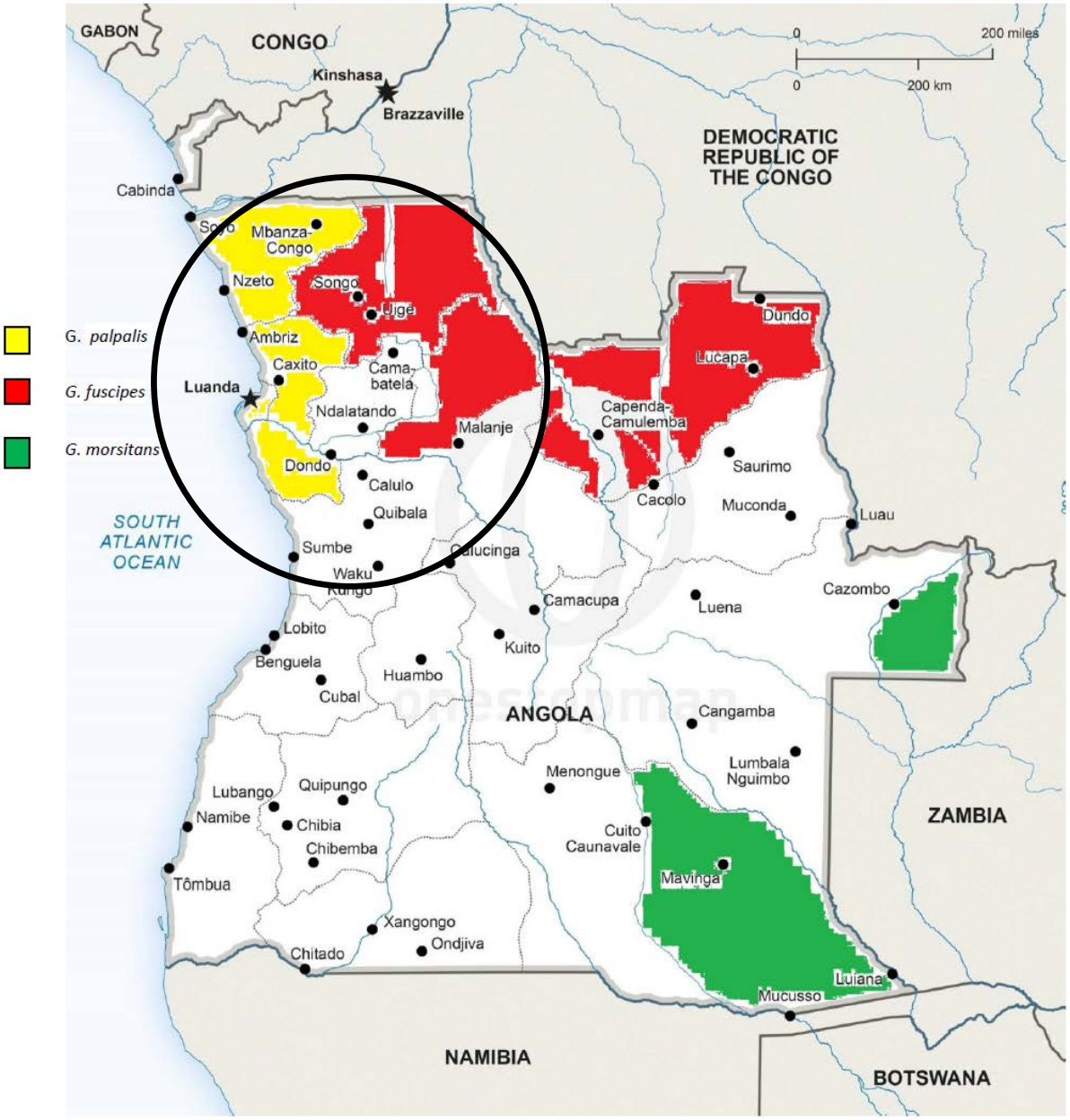


Figure 1-5. Geographical locations of tsetse fly sub-species in Angola, the circle highlights the provinces which are endemic for HAT. (Stanghellini & Josenando, 2001) and (Truc et al., 2011)

Angola is divided into 18 provinces (see Figure1-5). The chronic form of HAT, caused by *T.b.gambiense*, occurs in seven of the 14 provinces where tsetse flies are present (in Bengo, Cuanza Norte, Uige, Cuanza Sul, Zaire, Malange, and the periphery of Luanda). The acute form of the disease, caused by *T. b. rhodesiense* transmitted by *Glossina morsitans centralis*, could once be found in the southern province of Cuando Cubango (where 30% of new Angolan HAT cases were diagnosed between 1964 and 1974; Josenando, unpubl. obs.) (Truc *et al*., 2011).

1.6.2 Current situation of HAT

Although the current situation of the reported gambiense HAT number of cases is promising, but it should be observed with some caution. The infection affects mainly remote rural communities, where the health infrastructure is basic, and cases may be unrecognized or unreported. Some of the known active foci are difficult to access because they are insecure or the topography is challenging. When active case detection is conducted in previously inaccessible or neglected foci, gambiense HAT can be found, often at a substantial prevalence (Chappuis *et al*., 2010). Despite these gaps, knowledge of the current epidemiological situation has improved considerably in the past decade. In 1995, the gap between the numbers of reported and actual cases was estimated to be factor of 12 (WHO, 1998) by 2006 the number has decreased to factor of three (Weekly Epidemiological Record, 2006). Despite these achievements, much effort is still required to gather more accurate data on HAT in order that its designation as a Neglected Tropical Disease may be eliminated (Fèvre et al, 2008). Out of 36 countries listed as endemic, only 30 have received the exclusive support of WHO including assessing the epidemiological status of HAT or to establish control and surveillance activities, the 6 countries which have not received WHO support are; Botswana, Burundi, Ethiopia, Gambia, Namibia and Niger. (WHO, 2013). The support of WHO was given as; firstly technical assistance, which given by WHO staff. Secondly, provide the patients easy access to diagnosis; this is can be given by providing the equipment, reagents, logistics, and funds which will allow the national teams to reach HAT transmission areas to perform active case-finding surveys and set up passive surveillance. Also, WHO gave support in form of facilitating the access to the treatment; this includes the provision of drugs as well as patient accessibility (Simarro *et al*., 2011a).

Furthermore, between 2000 and 2010, 94 HAT cases were diagnosed in 19 non-endemic countries, of which 72% were due to *T. b. rhodesiense*: of them, 82% were diagnosed at the first stage. For comparison, among 26 cases of *T. b. gambiense* HAT cases, 77% were diagnosed at the second stage (Simarro *et al*., 2011b).

As for gambiense HAT, current trends in the number of cases reported by countries endemic for rhodesiense HAT are encouraging. Although, is difficult to make firm predictions about changes in the incidence of rhodesiense HAT. As long as reservoirs and environments for transmission exist, there will be a risk, and unpredictable ecological or social changes may result in extensive transmission (WHO, 2013).

http://apps.who.int/iris/bitstream/10665/95732/1/9789241209847_eng.pdf).

Finally, WHO have designed a road map to eliminate gambiense HAT by 2020, Elimination as a public health problem is defined as less than 1 new case per 10 000 inhabitants in at least 90% of foci, with fewer than 2000 cases reported annually at continental level. The 2020 goal is an intermediate step. The final goal, in accordance with the recommendations of the Neglected Tropical Disease (NTD), is to interrupt transmission of gambiense-HAT by 2030 (WHO, 2012).

In term of disease transmission risk, the risk is only limited to the foci of the disease and cannot be generalized on a national or continental scale. Based on new longitudinal analysis of the data included in the WHO Atlas of HAT and global population distribution layers, the extension and the location of different levels of risk has been estimated. Hence, it has been estimated that~70 million people live at different levels of risk for HAT infection. In addition to 57 million (81%) people distributed over an area of approximately 1.38 million km^2 are at risk of gambiense HAT; this population is distributed in 14 of the 24 countries listed as endemic for

gambiense HAT. Moreover, 12.3 million people are at risk of contracting rhodesiense HAT, and they are distributed over an area of 0.171 million km^2 (Simarro *et al*., 2012). More than 5 million people live in areas classified as high or very high risk for contracting HAT. The Democratic Republic of the Congo has the greatest number of people at risk (36 million) for HAT and the largest risk area (790,000 km^2) (Simarro *et al*., 2012) (Figure 1-6).

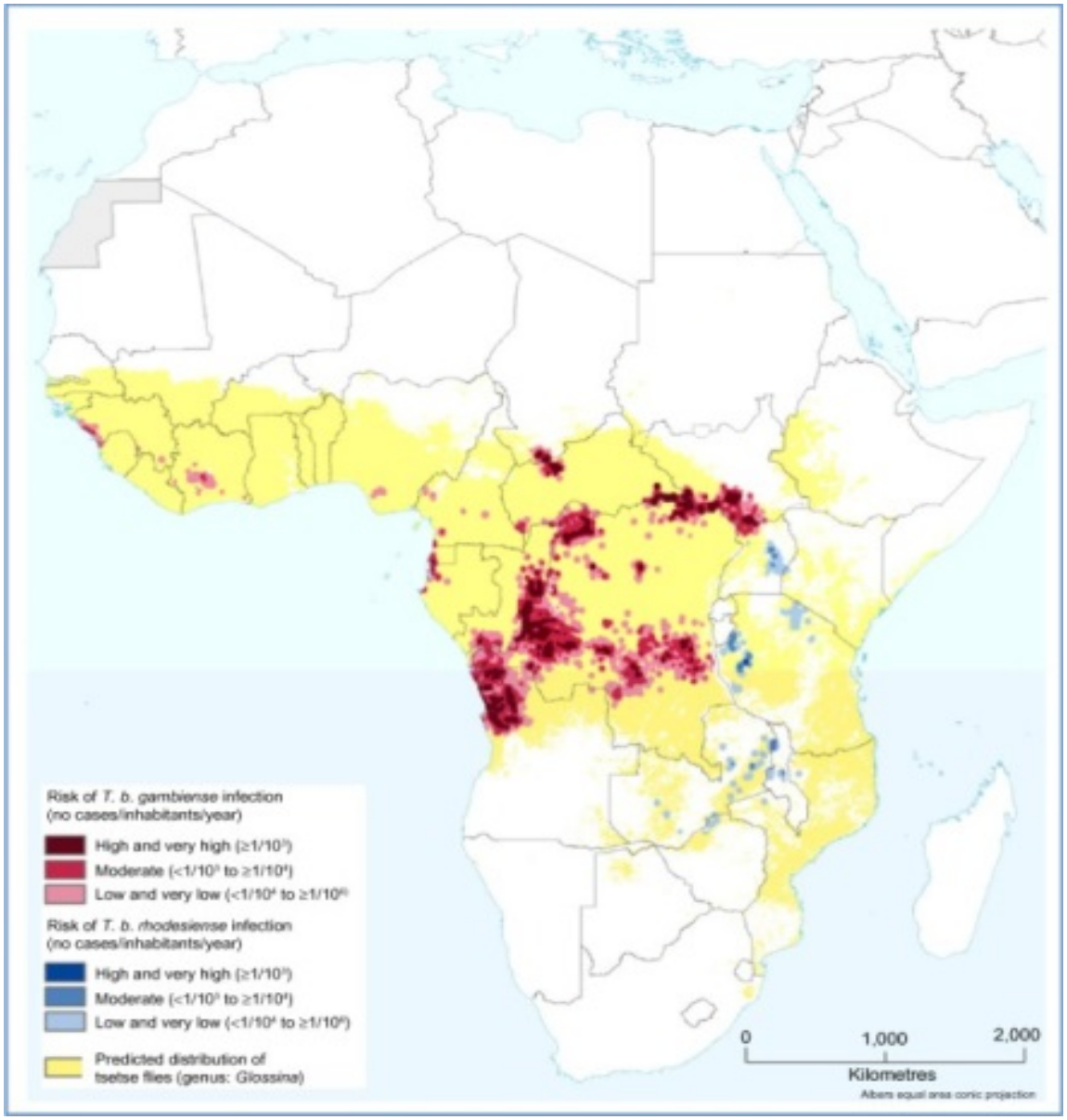


Figure 1-6. Population at risk of HAT in Africa

(Franco *et al*., 2014).

1.7 Diagnosis of Trypanosomiasis

Rapid and accurate diagnostic tools are vital in diagnosing HAT, also they play an

important role in meeting the “road map” set by WHO. The symptoms of the early stage of HAT including headaches, fever, fatigue and joint pain cannot considered as a definitive indicators of trypanosome infection, thus HAT is often mistakenly identified as other diseases such as influenza or malaria (Odiit *et al*, 2004)). Also, since the available drugs to treat sleeping sickness are either toxic or cumbersome to administer, and because symptoms are generally not specific enough to start treatment (Debaorggraeve and Büscher, 2010). The classical diagnosis of human African trypanosomiasis (HAT) is a 3-step approach: identification of suspects, parasitological confirmation, and disease staging. For the diagnosis of *T. b. gambiense* infection, relatively simple, reliable antibody tests can be used for screening (WHO, 2013). Most tests for detecting antibodies of *T. b. gambiense* contain selected VSGs of variable antigen types LiTat 1.3 and 1.5. Despite the abundance of VSG genes, most patients with *T. b. gambiense* infection appear to carry antibodies in their blood that react with at least one of the two VSGs.

For the diagnosis of *T. b. gambiense* infection, a rapid agglutination test is available that can be used in mass screening. Individual rapid diagnostic tests have also been developed, which are more appropriate for passive screening and are under evaluation. Tests based on immunofluorescence and enzyme-linked immunosorbent assays (ELISAs) are more appropriate for laboratory testing (WHO, 2013). The immune trypanolysis test is considered to be for use in reference laboratories. Serological tests detect antibodies only 3–4 weeks after infection, which may be one of the reasons for false-negative reactions. Cross-reactivity with other parasitoses may occur, particularly in whole blood or at low serum or plasma dilutions. In some rapid diagnostic tests for HIV and malaria, cross-reactions with HAT blood specimens may occur, decreasing the specificity of those tests (4, 5).

1.7.1 Parasitological Diagnosis

Microscopy is considered as the most widely used and simplest of diagnostic techniques and relies on direct observation (identification stage) of the parasites either in blood, lymph or cerebrospinal fluid. Although being widely used, these techniques are not considered sensitive enough to detect low parasite levels, especially in chronic cases (Luckins, 1992). The sensitivity of the microscopy examination can be increased; with haematocrit centrifugation technique (HCT) developed by Woo (1970) using centrifugation of microhaematocrit capillary tubes containing the blood sample and examination of the buffy coat/plasma junction under the microscope. In a positive diagnosis, trypanosomes are found wriggling at the junction of the buffy layer and the plasma (Woo, 1970), with an estimated detection threshold of 500 trypanosomes/ml (Chappuis *et al.*, 2005).

1.7.2 Card Agglutination Test for Trypanosomiasis

The card agglutination test for trypanosomiasis (CATT) is a rapid, simple assay for the detection of specific antibodies in patients with gambiense HAT (Magnus *et al*., 1978). Because of its simplicity, reliability and low cost, it is used in all control programmes for serological screening of populations at risk for *T. b. gambiense* (WHO, 2013).

The antigens originate from particular variable antigen types (VATs) of *T. b. gambiense* that are highly conserved across the range of this species and they are expressed early with infection. Therefore, the majority of infected individuals develop antibodies that cause visible agglutination when whole blood or serum is mixed with the antigen on the card (Eisler *et al*., 2004). The CATT antigen consists of complete bloodstream forms of T. b. gambiense variable antigen type LiTat 1.3. To prepare the antigen, trypanosomes are purified from infected rat blood, fixed, stained with Coomassie blue and freeze-dried. Kits contain the reagent, positive and negative

control sera and materials for performing the test on whole blood (capillary tubes, test cards, stirring rods, suction bulbs, syringe and droppers) (WHO, 2013).The specificity of the test for *T. b. gambiense* varies between 87% and 98%, with an average of 95% (Robays *et al*., 2004). Despite a specificity of about 97%, the positive predictive value remains limited when the test is used for mass screening in populations in which the overall prevalence of gambiense HAT is low (Checchi *et al*., 2011). On the other hand, False-negative CATT results may be obtained for patients infected with strains of trypanosomes that do not express the LiTat 1.3 gene, resulting in lower sensitivity of CATT in some endemic areas (Dukes *et al.*, 1992). False-negative CATT results in the presence of high antibody concentrations may also be found because of prozone (disequilibrium between the relative amounts of antibody and antigen) when undiluted blood, plasma or serum dilutions below 1/4 are tested (WHO, 2013).

However, decisions to treat should not be made when only CATT on undiluted blood is positive and parasite tests are negative. CATT titration on diluted blood, serum or plasma is imperative. Taking into account the disease prevalence and the feasibility of maintaining control activities, a cut-off point should be selected from a preceding study on serological characteristics in the area (Simarro *et al*., 1999).

Simarro et al. (1991) have previously reported a sera from 124 parasitologically confirmed individuals from the same focus that yielded a positive CATT/*T.b.gambiense* in 120 individuals (97%)and the end titre point was determined optimal at 1/8.. The authors suggested on the basis of these results, non-parasitologically confirmed CATT-positives should be divided into three groups: individuals with an end titre of 1/4 were considered not infected and were not followed-up; those with end-titres of 1/4 or 1/8 were suspected of trypanosomiasis

and were included in group A for follow-up andindividuals with end titres above 1/8 were strongly suspected of trypanosomiasis and were included in group B for follow-up. However, according to a CATT/*T.b.gambiense* study on 4753 individuals from the focus of Quiçama (Angola), it was concluded that the treatment should be given to all cases with an end titre > 1/8 even without parasitological confirmation, but always after lumbar puncture for staging and selection of the appropriate drug (Simarro et al, 1999). It should be kept in mind that some individuals could be treated unnecessarily, probablywith a not extremely toxic drug such as pentamidine, but many trypanosomiasis patients who otherwise would not have been detected would receive appropriate treatment.

The diagnostic method regarding the *T. b. rhodesiense* disease is different from that for *T. b. gambiense* in several ways; first there is no involvement of serological screening test for *T. b. rhodesiense*, the detection of suspected cases depends on the non-specific clinical presentation and history of exposure. Second, since the density of blood circulating parasites is higher than for *T. b. gambiense* parasitological confirmation is easier for *T. b. rhodesiense.* A thin or thick blood smear is usually sufficient to confirm diagnosis. Third, biological indicators such as haemoglobin and platelet counts and coagulation tests are more frequently or substantially changed in African trypanosomiasis caused by *T. b. rhodesiense* than in that caused by *T. b. gambiense*, but these findings remain non-specific (Brun *et al*., 2010).

1.7.3 Molecular Diagnosis

The identification and surveillance of infectious agents and in determining sources of infection had a major boost by the involvement of molecular tools.

Molecular identification is playing an important role in discriminating different

parasites with morphologically identical life cycle stages, include eggs or cysts, from faecal samples, or when attempting to match different life cycle stages of the same parasite from intermediate and definitive hosts (Crsione *et al.*, 2005). For instance, Thomson *et al.*, 2011 have reported that in some endemic areas of hookworm, human may be infected with more than one species. However, the identified eggs in the faeces are morphologically identical.

Also, molecular tools play a major role in studying the epidemiology of infectious agents (Traub *et al.*, 2005). According to Dorman (1992) the molecular definition is known as “a science that focuses on the contribution of potential genetic and environmental risk factors, identified at molecular level, to the aetiology, distribution and prevention of disease within families and cross populations”.

The most used tool is the polymerase chain reaction (PCR), which allows direct examination of clinical or environmental isolates has had an enormous impact on the genetic characterization, diagnosis and taxonomy of parasite (Lymbery and Thompson, 2012). PCR is considered as a sensitive and specific method for the detection and identification of human-infective trypanosomes .PCR also prevent the need for laboratory amplification of parasite isolates, as it was the major obstacle in characterising parasites refractory to in *vitro* culture (Lymbery and Thompson, 2012).

Defined gene sequences of infectious agents can be detected from small quantities of material and the resultant data can be detected by PCR, which can be used not only for diagnosis, but also to assess the effect of interventions on the population structure of infectious agents, assessment of intra-species diversity, and transmission studies (Lymbery and Thompson, 2012).

Furthermore, molecular tools play an important role in solving taxanomic issues that resulted in controversy in the past, when new species or strains were described on the basis of host occurrence, phenotypic characteristics and epidemiological

observations. Molecular applications function to solve such problems is can be described in two stages; first different genetic groups are found within what is ostensibly a single morphologically defined species, secondly these groups are defined as taxonomic categories, either at intraspecifc level or as different species or higher taxa (Lymbery and Thompson, 2012).

Moreover, the first development of PCR use in identification and detection of trypanosome was first published in 1989 by Moser *et al*. (1989). They used PCR to amplify a 188-base pair (bp) segment of the repetitive 195-bp nuclear DNA sequence of *T.cruzi* that is the most abundant sequence in this organism.

Kouadio et *al.* (2014) compared between buffy coat technique (BCT) and species-specific PCR to detect trypanosome species in 363 cattle samples from Côte d'Ivoire; out of 363 cattle examined with BCT, 33 were found positive with all trypanosome species accounting for an average of 9.09% prevalence whereas PCR using species-specific primers showed that 81 out of 363 cattle were infected with trypanosomes with an overall prevalence of 22.31%. *Trypanosoma congolense*

savanah type, *T.vivax* and *T. brucei sl.* accounted for 28.39%, 49.38% and 23.45% of the infection rate respectively.

1.7.3.1 Genetic Markers

There are 11 different pathogenic trypanosomes in trypanosomiasis endemic regions of Africa which are can be identified by molecular methods, and which relies on species-specific markers (Njiru *et al.*, 2005). In addition, the sensitivity and specificity of a diagnostic PCR is largely related to DNA sequence targeted by the primers. (Table 1-2 and Table 1-3) shows the number of markers designed for *T.b.gambiense* and *T.b.rhodesiense* respectively. Satellite DNA markers has been a common target used for detection and identification of *Trypanosoma* (*brucei spp., evansi* and *equiperdum*), the species *vivax* and *simiae*, and the three main *Trypanosoma congolense* types (savannah, forest, and Kenya Coast) (Clausen *et al*., 1998). In the case of satellite DNA one pair of primers is needed to detect each subgenus, species or type, which leads to the processing of three to five different PCRs per sample. Thus, many attempts were tried to carry out multiplex reactions, but the sensitivity of the tests decreased and some combinations of primers induced the synthesis of `non-specific (and non-expected) size products (Picozzi *et al.*, 2008).
Thus, the conserved sequences which are unique for the target group that occurs in multiple copies in the parasite genome, acts as templates in the diagnostic PCRs (Debaorggraeve and Büscher, 2010). Several DNA sequences have been used for the sensitive and specfic PCR-based detection of livestock trypanosome DNA in host blood and/or in tsetse flies (Masiga *et al*., 1992). *T. b. gambiense* has been characterized by another specific molecular marker, the *T. b. gambiense TgsGP*.

The sequence of the *TgsGP* which encodes a flagellar pocket glycoprotein, were found to be diagnostic for *T. b. gambiense* only (Berberof *et al*., 2001). Primers targeting this sequence were designed by Radwanska *et al*. (2002a) and were found to be specific for the amplification of *TgsGP* of *T. b. gambiense* when tested against non-*T.b.gambiense* populations revealing negative results. The detection limit of the reaction was evaluated; it reached 1000 trypanosomes/ml of blood after a single PCR.

Target	**Size of amplified Products**	**Primer Sequences**	**Ref**
AnTat 11.17 VSG gene	653 bp	AnTA: 5'-CAC AGA CGA CAG AAG CGA TA-3' AnTB: 5'-GAA AGT GGG AGT TGT TGCTC-3'	(Bromidge *et al*., 1993)
Non conserved sequence of kDNA	930 bp	MDF: 5'- CGC CCA TAA GAT TTC CGGTT-3' MDR: 5'- GGT GTA ATA CTC ACC CGGTT-3'	(Mathieu-Daude *et al*., 1994)
Non conserved sequence of kDNA	930 bp	MDF: 5'- CGC CCA TAA GAT TTC CGGTT-3' MDR: 5'- GGT GTA ATA CTC ACC CGGTT-3' MDN: 5'-CGA GGT ACT TCG AAA GGGTT-3'	(Schares and Mehlitz, 1996)
TgsGP	308 bp	Sense primers: 5'- GCT GCT GTG TTCGGAGAGC-3' Anti-sense primers: 5'- GCC ATC GTG CTTGCC GCT C-3'	(Radwanska *et al*., 2002a)

Table 1-2. The primers developed for use in *T. b. gambiense* molecular diagnosis

On the other hand, Moser *et al*. (1989) reported that *T. b. rhodesiense* universal primers that are used to prime the amplification of 177 bp sequences from Trypanozoon genomic DNA, are able to detect 0.1 pg of parasite DNA, this is about the amount of DNA calculated to be in a single trypanosome. On the other hand, the *SRA* gene marker which is found exclusively in *T.b.rhodesiense* allows the detection

of *T.b.rhodesiense* in reservoir hosts, the tsetse fly and humans (Nijru *et al.*, 2004). Nijru *et al.* (2004) concluded that *SRA* gene is highly valuable to detect and differentiate *T.b.rhodesiense* from other *T.brucei* in reservoir host, vector, and humans. Also, Wellburn *et al*. (2011) have reported the use of *SRA* gene as target for detecting *T.brucei*. According to the results of the study, the authors have suggested that *SRA* gene present a valuable marker to detect *T b rhodesiense,* in the domestic livestock reservoir.

However, *T.b.rhodesiense* infections within animal reservoirs are often chronic with very low levels of parasitaemias, which is largely based on the PCR test accuracy. A negative result may indicate the absence of SRA or simply indicate too little genomic material is present (Picozzi *et al.*, 2008). Thus, Picozzi *et al.* (2008) developed multiplex PCR to overcome this issue. This is a PCR reaction consist of two different sets of primers, therefore two different products will be amplified within the same reaction from the same starting material; one targeting the genus of parasite of interest, the second being sub species- specific. During each reaction, the PCR targets both *SRA*, and a single copy gene found within *T. brucei s.l.,* namely a phospholipase C (*GPI-PLC*). A successful amplification of *GPI-PLC* indicates enough *T. brucei s.l.* genomic material is present to detect a single copy gene.

Target	Size of amplified Products	Primer Sequences	Ref
SRA	743 bp	B537: 5'- CCA TGG CCT TTG ACG	Welburn *et al*.,
VSG	1121 bp	AAG AGC CCG-3'	2001)
		B538: 5'-CTCGAGTTTGCTTTT CTGTATTTTTCCC-3'	
	1.2 kb	SRA A: 5'-GAC AAC AAG TAC CTT GGC GC-3'	(Gibson, 2002)
		SRA B: 5'-CAG CAA CCA TAT TCA GAG CC-3'	
	284 bp	F: 5'-ATA GTG ACA AGA TGC GTA CTC AA CGC-3	(Radwanska *et al*., 2002b)
		R: 5'-AAT GTG TTC GAG TAC TTC GGT CAC GCT-3'	
	PLC: 239 bp	02: 5'- GGA GCC AAA ACC AGT	(Picozzi *et al*.,
	SRA: 669 bp	GGG CAC ATC-3'	2008)
	VSG: > 1 kb	03: 5'- AAG TAG CGC TGT CCT GTA GAC GCT TC-3'	
		651: 5'- GAA GAG CCC GTC AAG AAG GTT TG-3'	
		652: 5'- TTT TGA GCC TTC CAC AAG CTT GGG-3'	

Table 1-3. The genetic markers used in diagnosing *T.b.rhodesiense*

Moreover, Kabiri *et al.* (1999) developed PCR assay targeting expression-site-associated gene (*ESAG*) 6 and 7 DNA target sequences, to detect *T. b. gambiense* in blood samples. In *T. brucei*, the products of *ESAG* 6 and 7, located upstream of the variant surface glycoprotein gene in a polycistronic expression site. *ESAG*6 and *ESAG*7 are multicopy genes. The sensitivity of this assay was considerably high as it detected 87% of the 59 blood samples with specificity of 97%.

Furthermore, a development which plays an important role in the diagnosis of trypanosomes; using the length variation of multiple fragments within the 18S and 28S rRNA genes. Adams *et al*. (2008) have developed a new generic PCR method using fluorescent primers with the subsequent sizing of the products accurately and rapidly using an automated DNA sequencer. The authors have reported a comparison of the use of the new fluorescent fragment length bar-coding (FFLB) (Hamilton *et al*., 2007) and the using of ITS primers (below) (Adams *et al*., 2006) in examining midgut samples that were collected from Tanzania. The comparison revealed that using ITS primers, 78% of the examined flies were identified carrying a

parasite DNA of a different trypanosome species, while, using FFLB, 97% were found to be carrying trypanosome DNA. The authors suggested that the FFLB system is accurate, quick and has an extremely high level of identification of infection. Moreover, FFLB revealed the existence of a putative new species of trypanosomes that is most related to *T. brucei s.l.* by phylogenetic analysis. However, the costs of this technique can be the only obstacle since its requirement of expensive equipment and trained technical staff for application.

1.7.3.1 Internal Transcribed Spacer (ITS)-PCR

In trypanosomes, as in other eukaryotic organisms, the ribosomal RNA (rRNA) genes are transcribed as a large primary transcript, which is processed by stepwise endonucleolytic cleavage to yield the 18S, 5.8S and 28S units (Urakawa and Majiwa, 2001). At least five PCR reactions using species-specific primers would be required for each sample depending on the host species and the geographic distribution of the trypanosome species compared to the convenience of a single test.

Investigations have been carried to combine the species-specific primers in a multiplex reaction, but these trials have been shown to be less sensitive with some combinations of primers inducing the synthesis of non-specific and unexpected size products (Desquesnes *et al*., 2001). The most important obstacle of using species-specific primers is the lack of specific primers for all trypanosome species, leading to the inability of identifying new trypanosome species. For example, in a tsetse survey conducted in Tanzania, Lehane *et al*. (2000) failed to identify 13% of infections detected by microscopy using the available species-specific PCR. The development of the ITS region as a universal target for the diagnosis of different trypanosome species occurring in either the vertebrate host or the invertebrate vector, aimed to decrease the number of species-specific PCR reactions required for diagnosis and

allowed for the identification of new species infecting both the vector and vertebrate host.

The internal transcribed spacer (ITS) region of rDNA is a favourite target for a universal test because of its highly conserved flanking regions and size variability among trypanosomes species and subgroups. Also this locus has 100–200 copies and each transcribed unit is composed of 18S, 5.8S and 28S rRNA genes separated by two ITS regions (Njiru *et al.*, 2005; Hernandez *et al*., 1993). The two spacers, (ITS) 1 and 2 are known to vary in size between species and occasionally subspecies (figure 1- 5).

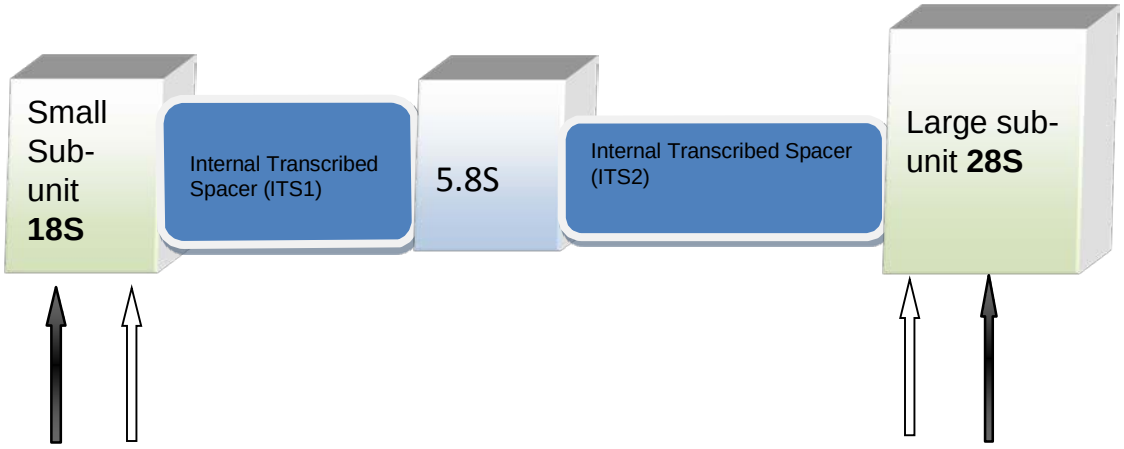


Figure 1-7. Generic structure of the ITS genes of eukaryotes. The white arrows show the annealing position for ITS3 (forward), within the 18S ribosomal small subunit and ITS4 (reverse), within the 5.8S ribosomal sub-unit. The black arrows represent outer primers, ITS1 and ITS2, designed to conserved regions within the 18S and 28S subunits (Cox *et al*, 2005).

Njiru *et al.* (2005) reported sensitivity of the ITS single reaction PCR as; *Trypanozoon* (95%) and *T. congolense* savannah (90.9%) and (77.4%) for T*. vivax* on a blood samples was collected from 373 cattle and 185 camels. On the other hand, Cox *et al.* (2005) developed a new nested PCR targeted to include both ITS, that was capable of detecting trypanosomes in the presence of host DNA and the PCR inhibitors present in blood. This method is very sensitive, it has the ability to detect DNA concentration at 49 pg ml^{-1}, which is considered as less than a single

trypanosome.

1.7.3.1 Mini- And Micro-Satellite Markers

Another marker which provides the greatest prospect is the mini- and micro-satellite markers (Hide and Tait, 2009). Microsatellites have only become popular as a genotyping tool, for instance *Plasmodium* (Wooden *et al*., 1993). This is largely related with the increasing number of genome sequencing projects (Traub *et al.*, 2005). Micro-and-mini –satellite are repeated sequences vary in length or sequence between trypanosome strains and can be amplified and analysed by population genetic methods (Hide and Tait, 2009).

In addition, PCR of microsatellite sequences is reported to be sensitive enough to detect trypanosome DNA in patients' biological fluids, and to detect mixed infections (Koffi *et al*., 2007). In addition, Biteau *et al.* (2000) managed to use microsatellite for distinction of *T.b.gambiense* group 1 from other subspecies. Simo *et al.* (2010) reported that by using seven microsatellite markers they were able to characterize *T.b.gambiense* isolates from Central Africa sub-region in order to improve knowledge on the population genetic structure of this subspecies. These markers confirmed the low genetic polymorphism within Central African *T. b.gambiense* isolates from the same focus and strong differentiation between different foci. Moreover, there is evidence from microsatellite genotypes and kinetoplast (mitochondrial) DNA sequences of 142 samples species-wide that *T. b. rhodesiense* is a host range variant of *T. b. brucei* (Gibson,2007; Balmer *et al*., 2007).

1.7.3.1 Random Amplified Polymorphic DNA (RAPD)

Another application of PCR in epidemiology is known as random amplified polymorphic DNA (RAPD). RAPD is a genetic fingerprinting tool for discriminating between strains that plays an important role in *T. b. rhodesiense* investigations. RADP uses random primers and low stringency conditions to amplify DNA and produce

fingerprint (Traub *et al.*, 2005).

This kind of approach is very useful with small amounts of DNA, which probably sampled the full extent of the genome rather than being restricted to certain genes or parts of the genome. However, the main disadvantage of this method is represented by it is highly susceptibility to contaminating host or other DNA – bands will be generated from any DNA template, it is not possible to interpret genetically and is unable to detect mixed infections of trypanosomes (Hide and Tait, 2009).

1.7.3.1 Amplified fragment length polymorphism (AFLP)

Many of the genetic typing assays have drawbacks in that they may require a relatively large amount of high-quality DNA or, as with RAPD technique, may be difficult to reproduce between laboratories (Jones *et al*., 1997). Many markers, especially isoenzymes, can evolve too fast for use as epidemiological markers, so that the result can be misinterpreted (Hide, 1999). Furthermore, in the majority of these methods, only a very limited part of the genome is covered through highly specific molecular targeting of one or more repetitive DNA elements. On the other hand, AFLP is an extremely useful and reliable technique for detecting polymorphisms and its reproducibility is reported to be very high (Vos *et al*., 1995). While the requirement for only small amounts of DNA makes it highly suitable for genotyping isolates on a large scale or for isolation of binary specific markers. Agbo *et al.* (2001) have used AFLP to investigate the genetic diversity of *T. brucei* isolates, and tested the technique for trypanosome genotyping in general. The authors reported that the variation between

Human-infective *T. brucei* subgenus is beyond only geographical origin, since the *T. b. rhodesiense* seem to be genetically more diverse, while the *T. b. gambiense* are more clonal or genetically more stable over time. Also, they have suggested that AFLP fingerprinting may aid the identification of polymorphisms linked to parasite factors of host tropism, and contribute to the understanding of host-parasite interactions at the molecular level.

1.7.3.2 Mobile Genetic Element PCR (MGE-PCR)

Mobile Genetic Element PCR (MGE-PCR) was developed to overcome the obstacle of contaminating host DNA (Hide and Tait, 2009). MGEs exist in a wide range of organisms from protozoa to humans. As genetic markers they have two crucial functions which are; repetitiveness and variability, which is largely related to improving the discriminatory sensitivity of molecular epidemiological markers (Hide and Tilley, 2001).

MGE-PCR is a single primer PCR technique that produces variable banding patterns using markers based on the trypanosomal mobile genetic RIME, which anneals to the 3′ end of the RIME sequence. RIME is a trypanozoon-specific mobile genetic element, present in the *T. brucei* genome at up to 400 copies (Tilley *et al.*, 2003). Furthermore, the MGE-PCR amplifies the element flanking regions through different stages of priming including; specific priming within the element, priming with adjacent elements, and mispriming in flanking regions (Hide and Tilley, 2001).

Tilley *et al.* (2003) used MGE-PCR on *T.brucei* stocks southeast Uganda, Kenya and Zambia, and they reported that MGE-PCR is highly specific to amplify trypanosomal DNA in the presence of host DNA, and detects high levels of variability within the *T. b. brucei* and *T. b. rhodesiense* groups and is therefore a powerful discriminatory tool for tracking individual *T. brucei* genotypes and strains.

Moreover, Nijru *et al.* (2011) used MGE-PCR in genotyping of *Trypanosoma evansi,* stocks that were isolated directly from camels and after their respective passage in mice were analyzed. Construction of a dendrogram using the MGE-PCR banding profiles revealed a clear distinction between *T. evansi* and *T. brucei*, as well as discriminating the *T. evansi* strains.

Also, MGE-PCR is useful for biological samples which are contaminated with host DNA, as Terry *et al.* (2001) have applied MGE-PCR *Toxoplasma gondii* which were directly isolated from sheep placental tissue. The MGE-PCR gel bands were absent in uninfected tissue. However, MGE-PCR gives some very complex multi-band patterns making the interpretation of result difficult, especially in mixed infection cases (Koffi *et al.*, 2007).

1.7.3.4 Loop-Mediated Isothermal Amplification (LAMP)

Residual co-amplification of irrelevant sequences still cause general drawback in nucleic acid amplification, especially for diagnostic use. Even though in the use of multiple primers such as nested PCR, which has improved amplification specificity for the target sequence. However, this consumes a lot of time, materials and reagents. Thus, a novel method has been developed called loop-mediated isothermal amplification (LAMP). LAMP has the ability to amplify a few copies of DNA to 10^9 copies in less than an hour under isothermal conditions 60- 65C^o with greater specificity (Notomi *et al*., 2000). LAMP utilizes four to six primers DNA polymerase *Bst* to specifically recognize six to eight regions of the target DNA sequence. Furthermore, the principle of LAMP is based on auto-cycling strand displacement DNA synthesis performed by DNA polymerase with high strand displacement activity and a set of two specially designed inner and two outer primers. An inner primer containing sequences of the sense and antisense strands of the target DNA initiates

LAMP(Notomi *et al*., 2000).

In addition, LAMP can be applied directly on tissue such as serum without extracting the DNA, also the visualization of the generated products can be done by the inclusion of fluorescent dyes for instance SYBR green (Adams and Hamilton, 2008). However, Wastling *et al*., (2010) have developed several simple endpoint detection methods have been developed for LAMP including; (i) visualization of turbidity; (ii) addition of hydroxynaphthol blue before incubation; (iii) addition of calcein with $MnCl_2$ before incubation and (iv) addition of Quant-iT PicoGreen after incubation. These four methods were applied to four LAMP assays for the detection of human African trypanosomiasis LAMP. The authors have concluded that hydroxynaphthol blue was the best of the compared methods for easy, inexpensive, accurate and reliable interpretation of LAMP assays for HAT. LAMP demonstrated to be a potential tool in the staging of HAT which is critical for therapeutic decisions (Namangala *et al*, 2012). Moreover, Nijru *et al.* (2008) have used LAMP to identify important members of subgenus *Trypanozoon.* They reported that the technique is not highly sensitive; but the sensitivity and reproducibility of the LAMP assay is indicated by the ability to detect the results visually without the need for sophisticated equipment, which indicates that the technique has strong potential for detection of HAT in clinical settings. In addition, Thekisoe *et al.* (2007) have used LAMP to identify human and animal trypanosomes, and they reported LAMP has the ability to detect as little as 1fg trypanosomes DNA, which is equivalent to ~0.01 trypanosome. They have concluded that LAMP has great potential of being used for diagnosis of trypanosomiasis in the laboratory and the field, particularly in countries that cannot afford modern molecular diagnostic techniques.

29

1.8 Treatment of HAT

The treatment for HAT is selected by first establishing the stage of infection; most of the affected countries use broad guidelines for the treatment of sleeping sickness as recommended by WHO (WHO, 2001). The chemotherapy of HAT has historical roots that date back to the pioneering work by Paul Ehrlich on dyes and arsenicals. Until recently, chemotherapy of HAT is given in form of three drugs; pentamidine, introduced in 1937, is currently available as pentamidine isethionate and is effective against early stage of the *T. b. gambiense* infection. Suramin which was introduced in 1916, is effective against early stages of both *T. b. gambiense* and *T. b. rhodesiense.* However, its use has been confined to *T. b. rhodesiense* infections. Melarsoprol a trivalent arsenical, was introduced in 1946 and until 1990, had been the only drug available for the treatment of late-stage *T. b. gambiense* and *T. b. rhodesiense* infections (Kuzoe, 1993), and eflornithine, was introduced in 1977, is the drug of choice for treatment of late-stage HAT caused by *T. b. gambiense*, but it is not recommended for *T. b. rhodesiense* (Fairlamb, 2003). (Figure 1-6) shows the different drugs involved in treatment of HAT and their chemical structures.

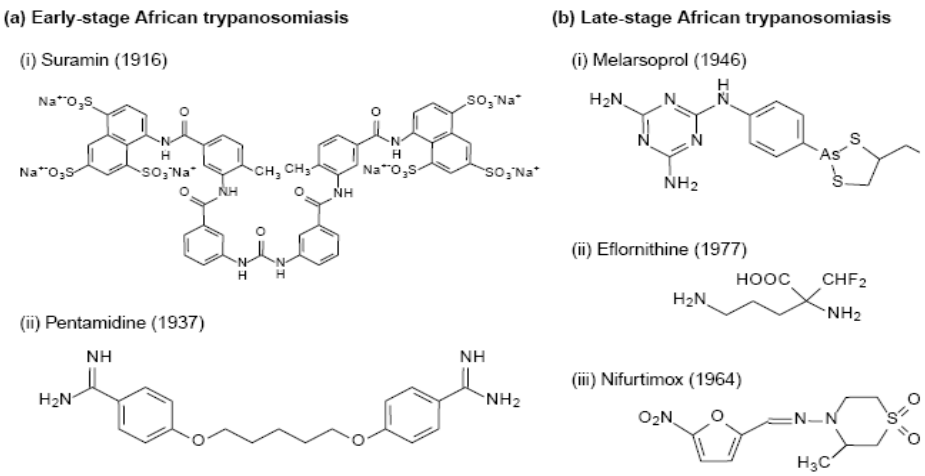


Figure 1-8. The chemical structures of drugs used in treatment of HAT. (a) Suramin and Pentamidine drugs used in the early stage of HAT, and (b) Melarsoprol, Eflornithine, and Nifurtimox used in the late stage of the HAT **(Fairlamb, 2003).**

Currently, according to Eperon *et al*. (2014) there are two anti-trypanosomiasis drugs that are still at clinical trial phase. Which are fexinidazole and SCYX-7158 for gambiense HAT. Should the trials record a success, fexinidazole and/or SCYX-7158 would be the first oral treatments to be used for both stage 1 and stage 2 sleeping sickness, thereby replacing the complicated diagnosis and treatment paradigm, which includes systematic lumbar punctures of every diagnosed patient to determine the stage of the disease before deciding which treatment to administer (DNDi, 2014). DNDi aims to deliver 2 new treatments that are safe, effective, easy- to-use (oral) and affordable for HAT stage 1 and 2 patients by 2018 (DNDi, 2014).

1.9 Aims

In this research study ways to improve the molecular field diagnosis of trypanosomiasis will be tackled. This study will be carried on 36 human patient blood samples collected on FTA card from Angola. The samples are positively confirmed for trypanosomiasis by CATT. Initially, nested ITS-PCR that was developed by Cox *et al*. (2005). The authors have suggested that testing a single punch from FTA card by ITS-PCR may lead to false negative, due to the uneven distribution of the parasite. Thus, alternative approach will be followed, which targeting more than a single punch for Chelex DNA elution prior the PCR amplification. To overcome the false negative limitation of the molecular tool in studying the epidemiology and diagnosis of sleeping sickness. Then, the use of different molecular tool (MGE-PCR) developed by Tilley *et al*. (2003) in the field diagnosis of HAT will be investigated on the eluted DNA from the Angolan samples, in interest of comparing the sensitivity and specificity between ITS-PCR and MGE-PCR. Finally, a gene family known as retrotransposon hot spot (*RHS*) genes/pseudogenes, that act as favourable target for mobile elements insertion in *T.brucei* (Brigaud *et al*., 2002), will be investigated through a serial of bioinformatical experiments. In order to find a potential marker to aid the epidemiology studies of HAT, and improve the prospects of MGE-PCR.

Chapter 2 **Materials and Methods**

To execute different types of studies different protocols and materials were followed and used, the diagram below (figure 2-1) illustrates the protocols that were used in this project.

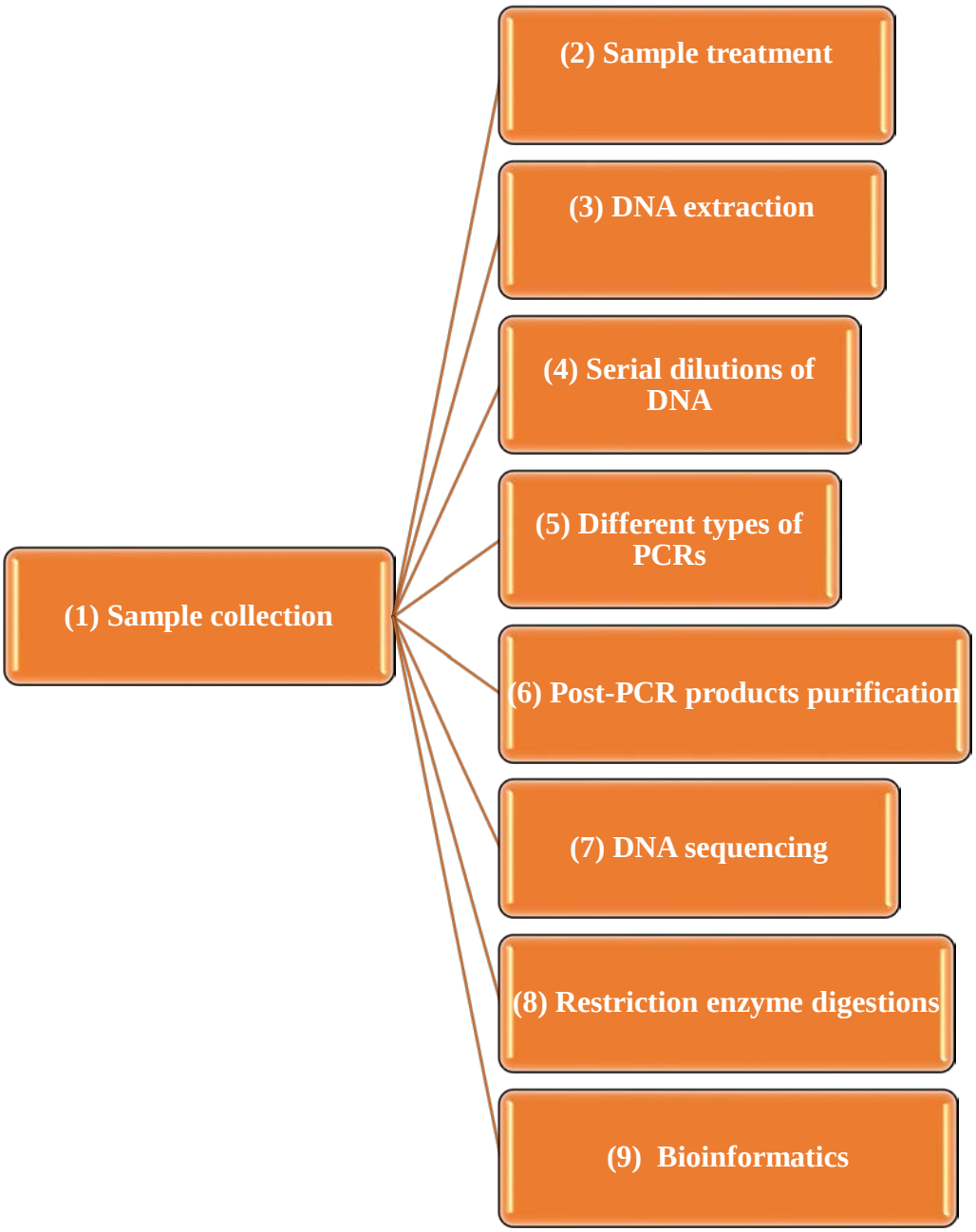


Figure 2-1. The methods that were followed in this study; the protocols are arranged in numerical order.

2.1 Sample collection

Thirty six blood samples were collected on FTA cards from different geographical parts of Angola, by the Sleeping Sickness hospital in Luanda, Angola in 2001. All these samples were infected with *T. brucei*; 21 were females and 16 were males and trypanosomiasis infection was confirmed by CATT. The blood of the patients was collected by a clinical official - Maria Giovetty. Blood samples from each patient were collected on FTA cards, preserved at room temperature and made available for the present study. By using a lancet the whole blood samples were taken from an auricular vein and then it was applied immediately in a spiral pattern on a Whatman FTA® card, as a part of routine diagnostic procedures. The cards were allowed to dry thoroughly at ambient temperature. Furthermore, Trypanosome genomic DNA (*T.b.rhodesiense*) in its liquid format will be used as positive control in the experiments, and to be diluted for sensitivity tests.

2.2 Sample Processing

Whole blood possesses potential PCR inhibitors including immunoglobulin G in plasma, haemoglobin in erythrocytes and lactoferrin in leukocytes (Al-Soud, 2000). Thus, it is crucial to treat the selected discs from the card in the following way prior DNA extraction and molecular investigation. Five 2.2mm discs were cut from the each spotted FTA® cards with a Harris Micro Punch ™ tool (Whatman Bioscience). Each disc was chosen randomly, but in preferential to the most spotted areas of the card and transferred to an eppendorf tube. After each time a punch was taken from the FTA card the Harris Micro Punch was cleaned on a filter paper to avoid any cross contamination. The selected discs from each sample were washed three times for 10 minutes in 200µl of FTA® purification reagent (Whatman Bioscience, Cambridge, UK) and rinsed twice for 5 minutes in 200µl of 1 mM Tris-EDTA buffer (Sigma Aldrich,

Dorset, UK) to remove traces of FTA® buffer. Discs were dried, preferably overnight.

2.3 DNA Chelex extraction

To carry out the designed molecular experiments the DNA was extracted from the dried discs by Chelex kit (Invitrogen). 5% Chelex® 100 suspension was used to extract the DNA from the discs; 5% Chelex® 100 suspension was prepared by dissolving 2.5g of Chelex® 100 in 50 ml of sterile water in 100ml autoclaved bottle, an autoclaved stirbar was placed inside the bottle then the bottle was placed on magnetic stirrer. Then, DNA was eluted by adding 60µl of 5% Chelex® 100 directly to the eppendorf tube that contains the treated disc, then heated for 30 minutes at 90Cº in a heat block.

2.4 Trypanosome DNA serial dilutions

In order to investigate the sensitivity of nested ITS-PCR the protocol was carried out on the different serial dilutions of trypanosomes DNA. The dilution serial was arranged from neat trypanosome DNA concentration of 80ng/µl to a dilution of 1:12. After each dilution the concentration of the DNA was measured by using Thermo Scientific 2000c nanodrop at 260nm.

2.5 PCR Amplification

Confirmed trypanosome genomic DNA (*T.b.rhodesiense*), in aqueous format, was used as the positive control in all PCR based approaches. The Chelex eluted DNA samples were tested by different molecular tools as described below.;

2.5.1 Nested ITS PCR

ITS-specific nested PCR was carried to detect the trypanosomes in the Angolan FTA card samples, and also it was carried on different concentrations of trypanosomes DNA to study the sensitivity of the protocol as described by Cox *et al*. (2005) by using the following primers. The outer primer sequences were ITS1 (5'-GAT TAC GTC CCT GCC ATT TG-3'), and ITS2 (5'-TTG TTC GCT ATC GGT CTT

CC-3') (MWG Biotech), and inner primer sequences ITS3 (5'-GGA AGC AAA AGT CGT AAC AAG G-3'), and ITS4 (5'-TGT TTT CTT TTC CTC CGC TG-3') (MWG Biotech) (Cox *et al*., 2005). PCR amplification was performed using a Robocycler (Stratagene). The reactions were carried in a volume of 25µl and included; 1.25U of Biotaq (Bioline), 2µM of each outer primer ITS1 and ITS2, 1µM of dNTPs and KCl buffer (Bioline) (final concentration of 10 mM Tris-HCl pH 9.0; 1.5 mM $MgCl_2$; 50mM KCl; 0.1% Triton X100; 0.01%(w/v) stabilizer). The reaction conditions were structured as; 1 cycle of 95 ºC for 7 minutes, followed by 35 cycles of 94 ºC for 1 minute, 55 ºC for 1 minute and 72 ºC for 2 minutes. Furthermore, 1µl of the first round was added to the second round reaction (nested PCR) which was placed in a fresh PCR tube and 24µl of the reaction mixture added. In the second round the outer primers (ITS1 and ITS2) were substituted by inner primers ITS3 and ITS4. The same reaction conditions used for the first round were applied in the second reaction.

2.5.2 MGE-PCR

MGE analysis was carried using a single primer (REV B) technique, through the analysis of the flanking regions of the element. The REV B sequence (5'CAG GGC TCT ACG GTG ATG GCT 3') anneals to the 3' end of the trypanosomal mobile genetic element RIME. The reaction was performed using a Robocycler (Stratagene). The reaction volume of 50 µl was used and contained 2 µM of REV B, 1 µM of dNTP mix containing 0.25 µM of each dNTP , 2U (5 U/µl) of Supertaq DNA polymerase, and 1.5mM $MgCl_2$ (Bioline) . The reaction conditions were as followed; 1 cycle of 94 ºC for 10 min, 10 cycles of 94 ºC for 45 s, 60 ºC for 1 min and 72 ºC for 2 min, followed by 25 cycles of 94 ºC for 45 s, 56 ºC for 1 min and 72 ºC for 2 min (Tilley *et al.*, 2003).

2.5.3 RIME A and RIME B Amplification

To study the of length variation of the RIME element, RIME A and RIME B primers were used (Tilly *et al*., 2003): RIME A (5′ AGC GTC TTG CTG GGA CAC CGT 3′) and RIME B (5′ AGC CAT CAC CGT AGA GCC CTG 3′). PCR amplification was performed using a Robocycler (Stratagene). Reaction volumes of 50 µl contained 2 µM of each primer, 1 µM of dNTP mix containing 0.25 µM of each dNTP, 2 U (5 U/µl) of *Taq* DNA polymerase (Bioline UK) and 1.5 mM $MgCl_2$. PCR cycles were as follows; a single cycle at 94 °C for 5 min, followed by 30 cycles at 94 °C for 45s, 55°C for 1 min and 72 °C for 2 min.

2.5.4 Mammalian specific tubulin PCR

This protocol is designed to target the mammalian tubulin gene. Tubulin genes are located in thousands of copies throughout the mammalian genome and the sequences of the primers are specific to mammalian tubulin only. This protocol was applied to verify if DNA was extracted from the FTA punches by Chelex extraction, as 18 samples from Angolan samples failed to amplify with the nested ITS-PCR. The PCR was carried out as described by Terry *et al.* (2001) using the following primer sequences: MtubF (5'- GTGAGTGCATCTCCATCCAT-3') and MtubR (5'-GCCCTCACCCACATACCAGTG-3'). Each PCR was carried out using 25µl volumes containing the following components: 10mM Tris-HCl pH 9.0, 1.5mM $MgCl_2$, 50mM KCl, 0.1% TritonX-100, 0.01% (w/v) stabilizer, , 2µM of each primer, 1mM total dNTPs and 1.25 Units of Biotaq (Bioline Ltd). The reaction conditions were as follows; one cycle of 94°C for five minutes, 40 cycles of 94°C for 50 seconds followed by 55 °C for 1 minute, 72°C for 1 minute 30 seconds with a final step of 72 °C for 10 minutes.

2.6 Visualization

A 10 µl aliquot of the final PCR products were mixed with 5µl orange loading dye (Thermo Fisher), and then were loaded into the wells of a 1.5% (w/v) agarose gel and electrophoresed at 110V for 1.5hr. The 1.5% (w/v) agarose gel was prepared by dissolving 1.5 g of agarose powder (Bioline, UK) in 100ml of 1xTBE buffer within a conical flask placed into a microwave oven at highest setting. The resulting agarose solution was then placed on a shaker to allow it to cool to approximately 50 ^{0}C and then 100µl of gel red (Biotium) was added to the mixture and it was poured into a gel casting tray with comb inserted. Following gel electrophoresis, DNA was observed using a ultraviolet light and images were captured.

2.7 DNA extraction from agarose gels

DNA was extracted from selected bands of the agarose gel using the PureLink® Quick Gel Extraction Kit (Invitrogen UK). After cutting the selected bands, they were transferred to an eppendorf tube and this was weighed to calculate the amount of buffer to be added to the tube. The protocol for purification is described in detail in the manufacturer's instructions (Invitrogen, UK). However, two important steps need to be taken into account; the band of interest was removed from the gel using a sterile blade to avoid any cross contamination and also, it is important to excise as little as possible of the agarose gel when cutting out the DNA band.

2.8 DNA sequencing of the PCR product

After successfully extracting the DNA from the gel as described above, the DNA concentration and purity were measured using a Nanodrop (2000/c, Thermo Scientific) at 260/280nm absorbance ratio. Before sending the samples to the sequencing company (Source Bioscience, Nottingham), the samples were calibrated according to the requirements specified by Source Bioscience . The samples were

sent by special delivery post to avoid any potential DNA, or primer degradation. The resulting DNA data was analysed by Finch TV.

2.9 Restriction enzyme digestion

Due to the complication related with the sequencing of REV B amplified products, restriction enzyme digestion was carried to cut the amplified products. The digestion was carried out in a 20µl reaction as follows: 7 µl DNA, 1.5 unit restriction enzyme (Mbo I and AlwNI, NEB UK), 2 µl 10 x digestion buffer and 9.5 µl water. Initially the MboI reaction was incubated at 37Cº for 2 hours, then it was incubated overnight; however, it failed to digest the PCR products. Consequently, AlwNI was added to the and then reactions were incubated overnight.

2.10 Bioinformatics and analysis of the *T.brucei* genome TREU927

Due to the challenges with MGE-PCR a further study into MGEs was required since retrotransposon hot spot genes (*RHS*) are favourable target for mobile elements insertion including RIME. Thus, they were targeted in this study in order to understand the behaviour of the mobile elements and to assist development of a diagnostic marker.

2.10.1 Data source

The *T.b.brucei* TREU 927 strain was used in this study and the *RHS* genes of TREU 927 sequences were accessed through the NCBI database http://www.ncbi.nlm.nih.gov/mapview/) and also tritrypdb (http://tritrypdb.org/tritrypdb/). Two terms were used in the search for *RHS* genes in *T.b.brucei; "RHS"* and *"retrotransposon hot spot".* Amino acid sequences of *RHS* genes were retrieved from GeneDB (http://www.genedb.org/Homepage). In addition,

RHS nucleotide sequences were also extracted. The insertion *ingi*/RIME was investigated in more detail and hence mobile element RIME and *ingi* sequences

were obtained from NCBI (http://www.ncbi.nlm.nih.gov/nuccore) under the accession numbers K01801.1 and JQ917146.1 respectively. These were subsequently used as reference sequences to screen for RIMEs inserted in the obtained RHS genes.

2.10.2 Multiple sequence alignment

The retrieved amino acid and nucleotide sequences of *RHS* genes were aligned with RIME sequences. Multi- alignment of DNA sequences or protein sequences were performed online using the ClustalW2 programme (Multiple Sequence Alignment tool, version 2.1) from the EMBL-EBI web server (http://www.ebi.ac.uk/Tools/msa/clustalw2/) with the default settings (Larkin *et al*., 2007). Also, Mview from EBI (http://www.ebi.ac.uk/Tools/msa/mview/) (McWilliam *et al*., 2013) was used to study the multiple sequence alignments.

2.10.4 Bioinformatics software

Various computer software, including molecular evolutionary genetics analysis version 5.0 (MEGA 5) (Tamura *et al*., 2011), was used to construct phylogenetic trees. Alignments were manually edited, and regions containing more than 50% gaps, or those were low sequence conservation precluded accurate alignment, were removed. Support for the inferred topology was estimated from 100 bootstraps. Snap Gene (2.4.3) was used for sequence viewing and extracting schematic diagrams, and nonsynonymous and synonymous substitution rates (denoted as Ka and Ks, respectively) were estimated using the KaKs Calculator (1.2) (default setting) (Zhang *et al.,* 2006).

Chapter 3 Identification of African trypanosome species from FTA card using ITS-PCR

3.1 Introduction

The control of sleeping sickness is highly related to our knowledge of the epidemiology of the disease, which is based on highly specific and sensitive screening and surveying methods (Hutchinson *et al*., 2003). Sleeping sickness in Angola is currently on the decrease as less than 100 cases been reported in 2015 (WHO, 2015). The full background of sleeping sickness in Angola is mentioned in (Chapter 1).

PCR has the ability to detect the presence of DNA sequences specific to the organism in question. The advantage of this technique is that it can specifically amplify a very small amount of parasite material which will help in identifying low parasitaemias with high sensitivity. For example, Masake *et al.* (2002) reported a successful detection of one trypanosome per 10 ml of cattle blood.

Three different types of PCR are used for the diagnosis of trypanosomes, including conventional, multiplex and nested reactions. Conventional PCR reactions mainly depend on the use of primer sets targeting a specific DNA sequence. In multiplex reactions, more than one primer set can be used which allows more than one DNA sequence to be targeted. Finally, nested PCR depends on the running of two separate reactions where the amplification product from the first reaction is used as a template for the second reaction (e.g. Cox *et al*., 2005).

However, in some cases, PCR assays have been linked to sampling issues, including samples that are mouse-passaged, and thus mouse adapted, a process which can cause a lowering of the epidemiological studies. In these cases some of the parasite isolates do *not* survive (Masiga *et al*., 1992) leading to a loss of species

or strains, and a selection and sampling bias (Welburn *et al*., 2001).

Thus, breakthroughs in matrices designed to improve sample collection and archiving such as the FTA card have been introduced. FTA cards have been designed to fix and store nucleic acids directly from tissues, allowing the collection and archiving of nucleic acids (Belgrader *et al.*, 1995; Picard-Meyer *et al.*, 2007). The FTA matrix is infused with protein denaturants that cause lysis of cells and any organisms on contact. In addition, these chemicals have the ability to inhibit saprophytes when drying to ensure the safe handling of cards without risk of biohazard (Belgrader *et al*., 1995; Rogers and Burgoyne, 1997). FTA technology also includes chelating agents and a free-radical trap that is designed to deal with atmospheric pollutants, therein protecting the trapped nucleic acids for at least six years at room temperature (Rogers and Burgoyne, 1997). Therefore, the use of FTA cards has been extended to include DNA detection from pathogenic protozoa and pathogenic organisms isolated from foods and clinical specimens using PCR (Orlandi and Lampel, 2000). FTA cards have been used for blood storage (Devost and Choy, 2000; and Gutierrez-corchero *et al.*, 2002), as well as in detecting bacterial pathogens (Lampel *et al*., 2000; Rogers and Burgoyne, 1997), the detection of free living protozoa in water samples (Hide *et al.* 2003), the detection of viral genomes (Owor *et al*., 2007; Picard-Meyer *et al*., 2007) and in forensic human biology (Hsiao *et al*., 1999).

Furthermore, in trypanosome diagnosis, FTA cards have been used for the storage of many materials including blood, buffy-coat and extracted DNA for the detection of trypanosomes using PCR (Becker *et al*., 2004; Cox *et al*., 2005). FTA cards have aided the diagnosis process in terms of time, storage and in field diagnostic laboratories. Adams *et al*. (2006) have reported the use of FTA cards for storage of midgut material originating from tsetse flies. In their research they noticed that the detection of trypanosomes using PCR was lower than expected (56.9%), due to various possible factors including the sub-optimum binding of trypanosome DNA to the FTA cards, and the uneven binding of the trypanosome DNA throughout the card due to competition with midgut material. However, the results have been improved, reaching 78% after storage of the midgut samples in 100% ethanol and purification of DNA using ammonium acetate precipitation.

In conclusion, the advantages of using FTA card technology rather than other methods include; lower sample volume (100 µl), safer handling, easier immobilisation of the DNA, less storage space being needed, ease of transportation and reduced financial costs (Gutierrez-corchero *et al*., 2002). However, there are various obstacles in terms of using FTA cards including the notable possibility that the punch that is punched from the FTA card may contain a small volume of the material estimated to be 1% of the total blood sample captured on the FTA card (Ahmed *et al*., 2011). For PCR based applications the minimum required volume to be assayed is typically only a single microlitre. Based on this there is a possibility that the punch samples taken from FTA cards for PCR-testing may result in an underestimation of the prevalence in the host population, particularly when that population comprises individuals with very low parasitaemias (Becker *et al*., 2004).

In addition, Chelex 100 is known as a chelating resin that possesses a high affinity for polyvalent metal ions. The Chelex resin consists of styrene divinylbenzene copolymers containing paired iminodiacetate ions, which function as chelating groups. Singer-Sam *et al.* (1989) have suggested that the presence of Chelex in high temperatures prevents the degradation of DNA by chelating metal ions that may act as catalysts in the breakdown of DNA at high temperatures in low ionic strength solutions. In addition, Singer-Sam *et al.* (1989) state that the use of Chelex 100 has led to an increase in the signal from the PCR amplification of small amounts of DNA released from small numbers of tissue culture cells that have been boiled.

The protocol is simple, rapid, involves no organic solvents and does not require multiple tube transfers for most types of samples; these factors reduce the chances of cross contamination. The use of Chelex DNA in the extraction of semen and very small bloodstains using Chelex 100 has proven to be as efficient as or more efficient than using proteinase K and phenol-chloroform extraction. The DNA extracted from bloodstains has been found to contain less PCR inhibitors when prepared using the Chelex method (Walsh *et al*., 1991).

In this chapter the internal transcribed spacer nested PCR (ITS-PCR) will be tested on extracted DNA from blood samples spotted onto the FTA cards by Chelex extraction. ITS –PCR is a method which detects the inter-specific length variation of the ITS regions of ribosomal genes producing a unique size of PCR product for each species - as reported by Cox *et al*. (2005). ITS region of rDNA is a favourite target for a universal test because of its highly conserved flanking regions and its size variability among trypanosomes species and subgroups. Also this locus has 100–200 copies and each transcribed unit is composed of 18S, 5.8S and 28S rRNA

genes separated by two ITS regions (Hernandez *et al.,* 1993; Njiru *et al.*, 2005). The two spacers, (ITS) 1 and 2 are known to vary in size between species and occasionally subspecies (figure 1-5).

The main advantage of using nested ITS-PCR is its ability to identify and differentiate all clinically relevant African trypanosome species. The high copy number located within inter-species variation in the length of the ITS region allows for identification and, importantly, species differentiation of trypanosomes (Desquesnes *et al.*, 2001; Cox *et al.*, 2005). This is in contrast to several molecular *Trypanosoma* identification methods, such as minisatellite markers, microsatellites or RAPD. These were outlined in detail in Chapter 1.

3.1.1 Aims and Objectives

The aim of this chapter is to investigate the use of specific molecular diagnostic tools on a set of diagnosed *Trypanosome brucei gambiense* patient samples, collected on FTA cards, from Angola. The detailed objectives are as follows:

1. To conduct an analysis of the demographics and characteristics of the group of Angolan sleeping sickness patients used in this study
2. To investigate and validate the use of ITS-PCR as a diagnostic tool for the detection of infection in the Angolan patients.
3. To investigate the use of Chelex extraction in improving and optimising the performance of the ITS-PCR on the samples collected from the Angolan patients.
4. To investigate the relationship between the performance of the ITS-PCR and the demographics and known infection status of the samples collected from the Angolan patients.

3.2 Methods

In this chapter one of the molecular diagnoses of HAT protocols (chapter 2) will be carried out on thirty six blood samples taken from patients who have tested positively for trypanosomiaisis by CATT collected on FTA cards from Angola. After the DNA elution from five punches from each FTA card by Chelex, the samples underwent a mammalian tubulin PCR was carried out to confirm successful DNA elution and suitability for PCR. Detection of trypanosomes was carried out using nested ITS- PCR. The correlation between the performance of the ITS-PCR and the demographics and known infection status of the samples collected from the Angolan patients will be studied using statistical tests.

3.3 Results

To investigate the usefulness of FTA cards as a method of collecting, storing and diagnosing human sleeping sickness, a collection of samples from sleeping sickness patients in Angola was studied. Specifically, the aim was to evaluate the role of Chelex eluted DNA as a method of detecting and differentiating trypanosome species from blood samples spotted onto FTA cards. Blood samples from thirty six parasitologically confirmed patients, were tested using a nested ITS PCR which amplified the variable ITS region of the ribosomal genes.

3.3.1 Description of the samples of infected blood immobilised on FTA cards from human trypanosomiasis patients in Angola

Blood from thirty-six human patients all of which were infected with *Trypanosoma brucei gambiense* were collected FTA cards using blood that was taken for diagnostic purposes. These patients were all confirmed as testing positive for sleeping sickness using the Card Agglutination Test for Trypanosomiasis (CATT) and also microscopic analysis of a lymph node aspirate (Lymph node puncture, LNP) (see Appendix 1). Some patients were also confirmed as testing positive for Lumbar Puncture (Appendix 1), known as "liquide céphalo-rachidien" (see Appendix 2 -LCR – Cerebro-spinal fluid aspiration) and also for CATT Titration (Appendix 1, TIT). Data on previous treatments, the location of patients and other demographic profile factors were obtained (see Appendix 1).

(Figure 3-1) shows the locations where the patients lived at the time of the blood samples being taken and (Table 3-1) is a summary of the sampled patients and other relevant profile information. (Full details can be found in Appendix 1.) As can be seen from the map, the patients sampled were resident in different regions of Angola, and the number of collected samples varied from region to region depending on who presented themselves to the clinic on the days the researcher was present. A significant majority

of the patients came from Cuanza Norte.

Figure 3-1. Angolan map showing the locations of the samples. Locations are represented by red stars. Scale: 1cm=100mile. The map was generated by National Geography Map Maker (http://mapmaker.education.nationalgeographic.com/?ar_a=1&b=1&ls=000000000000)

Patients number	Age/Gender	Pre-treat (3 days)	Treatment (days)	Vector (Cattle/testse)	Location
1	20/female	None	None	Present	**Luanda**
2	27/female	Yes	Pentamidine	Present	**Zanga**
3	45/female	Yes	Melarsoprol (8)	Present	**Uige**
4	25/female	Yes	Melarsoprol (8)	Present	**Uige**
5	20/female	Yes	Melarsoprol	Present	**Malange**
6	42/female	Yes	Pentamidine	Present	**Malange**
7	38/female	Yes	Melarsoprol	Present	**Kuando**
8	17/female	None	None	Not Present	**Uige**
9	25/female	Yes	None	Present	**Gulungo**
10	22/female	None	Pentamidine	Not Present	**Lukala**
11	12/female	Yes	Melarsoprol (1)	Present	**Bengo**
12	71/female	Yes	Melarsoprol	Present	**K-Norto**
13	21/female	Yes	Melarsoprol (8)	Not present	**Kinxassa**
14	50/female	Yes	Melarsoprol	Not Present	**Uige**
15	40/female	Yes	Melarsoprol	Present	**Malange**
16	9/female	Yes	Melarsoprol	Not Present	**Malange**
17	38/female	Yes	Melarsoprol (6)	Present	**Malange**
18	25/female	Yes	Melarsoprol	Present	**Bengo**
19	32/female	Yes	Melarsoprol	Present	**Bengo**
20	10/male	Yes	Melarsoprol	Present	**K-Norte**
21	13/male	Yes	Melarsoprol (7)	Present	**Zaire**
22	23/male	Yes	Melarsoprol (2)	Present	**Bengo**
23	25/male	Yes	Melarsoprol (6)	Present	**K-Norte**
24	12/male	Yes	Melarsoprol	Present	**K-Norte**
25	33/male	Yes	Melarsoprol	Present	**Bengo**
26	35/male	Yes	Melarsoprol (2)	Present	**Malange**
27	45/male	Yes	Pentamidine	Present	**Bengo**
28	15/male	Yes	Melarsoprol (4)	Present	**Malange**
29	26/male	Yes	Melarsoprol (4)	Present	**Huambo**
30	42/male	yes	Melarsoprol (4)	Present	**Moxico**
31	32/male	Yes	Melarsoprol (3)	Present	**K-Norte**
32	32/male	None	None	Present	**Uiege**
33	21/male	None	None	Present	**K-Norte**
34	32/male	None	None	Present	**K-Norte**
35	15/female	None	None	Present	**Bengo**
36	**47/male**	**None**	**None**	**Present**	**Bie**

Table 3-1. Demographic data for the Angolan patients. This table shows the demographic details of patients' blood collected on the FTA cards.

3.3.1 Analysis of the patient collection

When the blood samples were collected a range of patient profile information was recorded, including the following information: gender, age, family medical history, and place of residence, recent geographical movements and any mobility/disability issues the patients may experience (see Appendix 1). To facilitate a valid interpretation of the final outcome of the present study, it was important to obtain a comprehensive understanding of the patients' profiles in order to compare these variables with the DNA diagnostic data (see subsequent section).

Several key demographic features are worthy of note. Firstly, the majority of the patients were females - out of the 36 collected samples, 20 were from females. Moreover, the ages of the patients varied, with a reasonably spread out representation of each age group occurring and the youngest patient being nine years old. (Figure 3-2) shows the ages of the samples according to age-range groups in a pie chart.

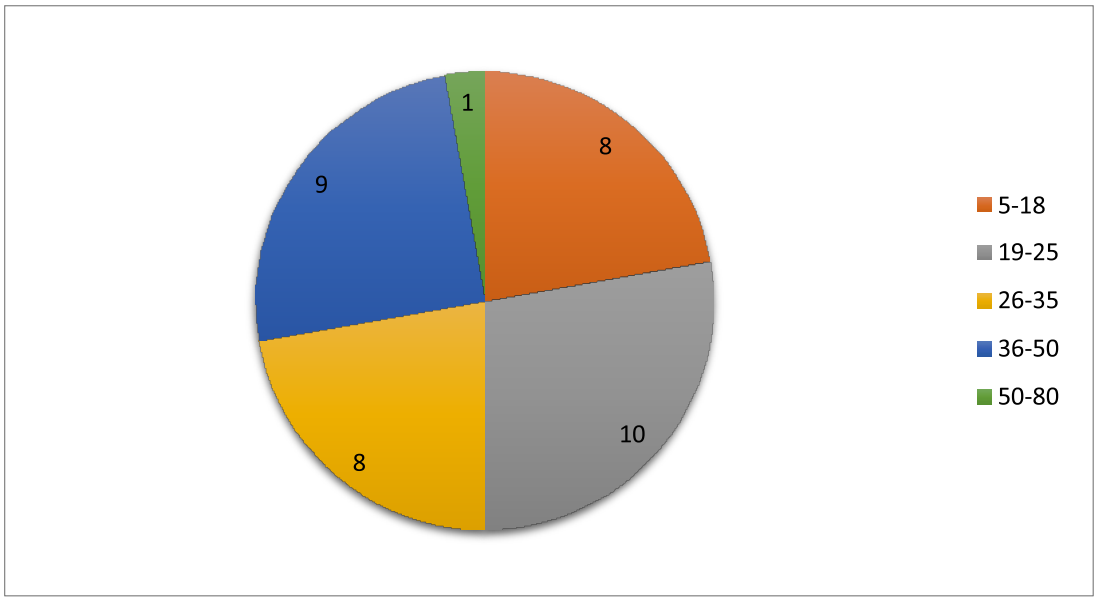


Figure 3-2. The age groups of the Angolan patients who participated in the study

According to (Figure 3-2) the majority of the cases were represented by 19-25 age group and the least was represented by the elderly (50-80 age group).Patient no. 16 who was nine years old, to patient no. 12 who was 71 years old at the time the blood samples were taken (62 years) (Table 3-1).

In addition, the samples were analysed according to medical history including familial history and also patients' past receipt of medical treatment for sleeping sickness. (Figure 3-3) shows that 63.8% of the patients had no family history of sleeping sickness, whereas 13.8% had family members who had experienced sleeping sickness and 19.3% had received previous treatment having been diagnosed with sleeping sickness

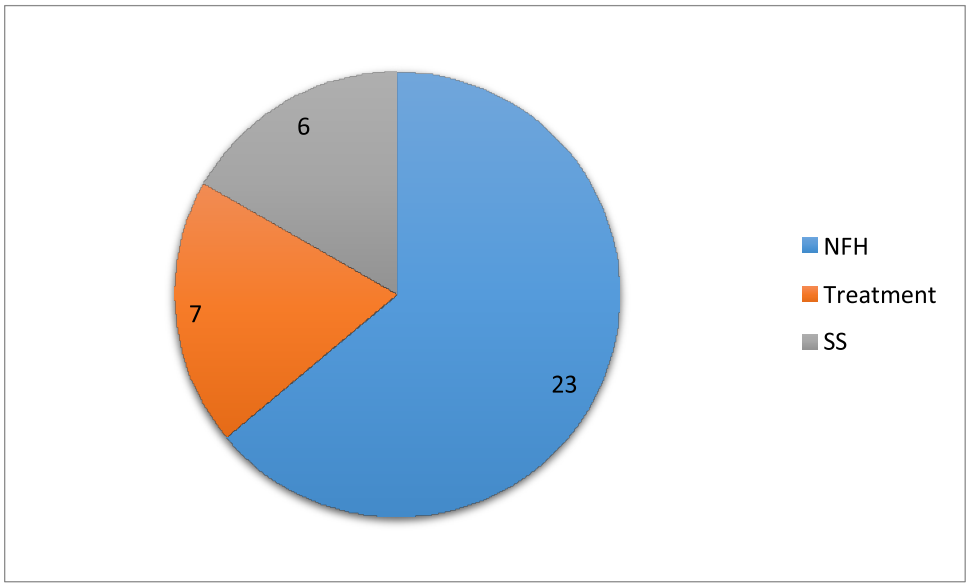


Figure 3-3. Angolan patient medical history. The number of cases that have sleeping sickness (SS) in the family, that have no family history (NFH) and the patients that in the past have received treatment against sleeping sickness

.

Although patient nos. 26, 29, 30 and 31 did *not* report having a family history and nor had they been treated in the past, they *have* travelled to different provinces in Angola as part of their work in the army.

Finally, the period during which the patients reported experiencing symptoms ranged from just a few months to a matter of years, as shown in (Figure 3-4). One of the longest periods of symptom experience reported, which stands out in the sample overall, was reported by patient no. 8 who had experienced symptoms for over seven years.

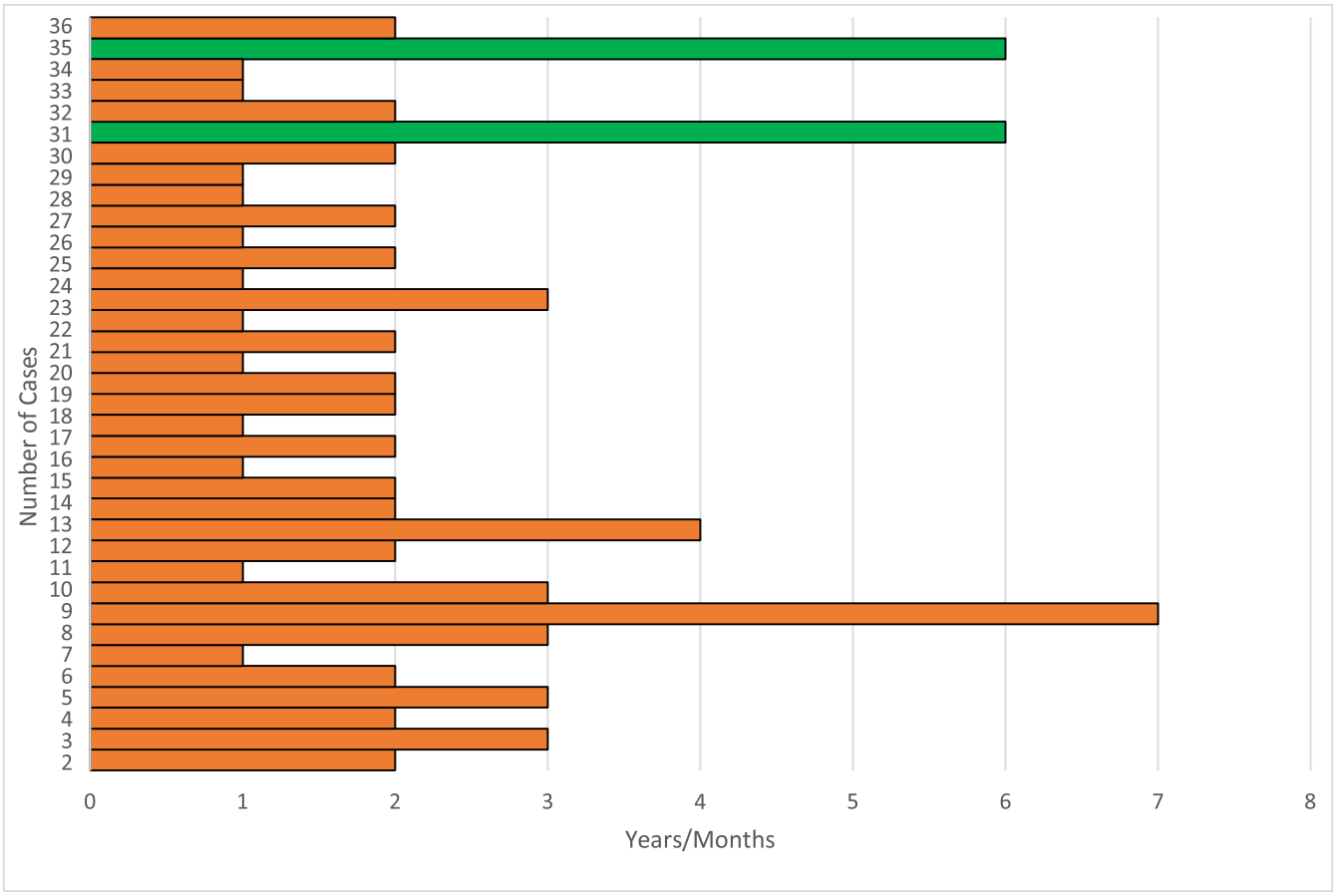


Figure 3-4. Duration of symptoms for each patient. The duration of symptoms for all patients is presented in years, except for patients 31 and 35 (marked by green colour) whose symptomatic duration is measured in months.

To summarise, the samples were collected and stored on FTA cards. The samples were analysed according to the available information on the patients as outlined above. The analysis of the sample showed that the patients were resident in different parts of Angola, with the majority of the patients being female. The oldest patient was a female of 71 years old (patient no. 12) but children were also included in the sample with the youngest being just 9 years old (patient no. 16). The sample showed a reasonable degree of variation in terms of the duration period in which symptoms had occurred before the collection of blood took place in the clinic. Finally, the patients showed a reasonable degree of variation in the duration period of symptoms before the collection of the blood. Patient no. 9 showed the longest symptomatic

period (seven years) with the shortest period being for just six months (patient 31 and 35). It is not clear why there was such variation but it might be simply that it was difficult for some patients to get to the hospital as Angola was in the throes of a civil war at the time. For example, patient 31 resides in Uige which is 210 miles from the sleeping sickness hospital in Luanda, and this may likely have affected her ability to turn up to the screening.

3.3.2 DNA amplification and diagnosis of trypanosomiasis

In order to effectively detect the *Trypanosoma*, the application of the nested ITS-PCR on DNA eluted from field samples and the efficacy of the Chelex protocol will be examined.

3.3.2.1 Sensitivity of nested ITS-PCR

As far as is currently known, the sensitivity of the nested ITS-PCR has not been tested on Chelex eluted DNA from FTA card punches. However, the findings of Cox *et al*. (2005) which indicate that the PCR amplification from a single punch may lead to false results will be tested. Nested ITS-PCR was carried out on a single punch from FTA cards spotted with neat known trypanosome DNA and blood from patients who tested positive for trypanosomiasis by CATT collected from Angola. The ITS primers failed to amplify all the samples except the positive control (liquid trypanosome DNA) (see Figure 3-5).

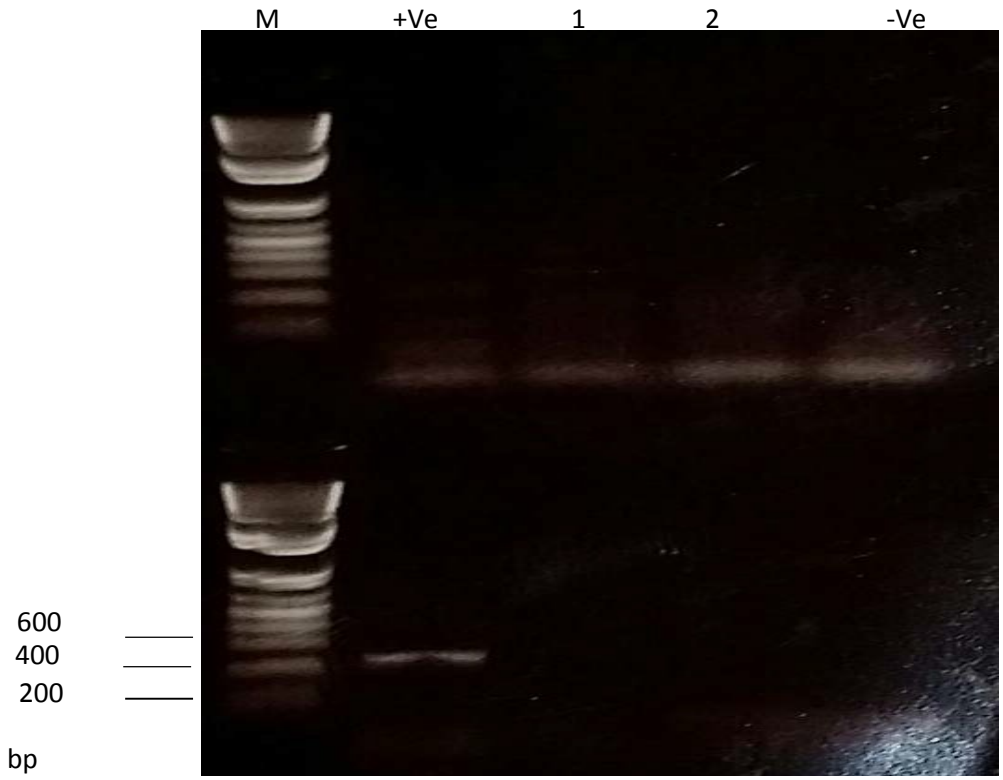


Figure 3-5 Nested ITS-PCR of trypanosome DNA and blood spotted on FTA cards. The upper and lower gel images represent the first and second reactions of nested ITS-PCR respectively. (M) Represent the 1kb DNA marker (Bioline). (+Ve) indicates the positive control (aqueous trypanosome DNA).Lanes 1 and 2 indicate the single punch taken from FTA card spotted with known trypanosome DNA and the blood of a patient who positively tested for trypanosomiasis respectively. (-Ve) is the negative control (water).

From (Figure 3-5), the sensitivity of the nested ITS-PCR can be seen to be 0%. However, it is worth noting the advantage of using nested ITS-PCR rather than a single round PCR, as the assay failed to amplify all the samples including the positive control in the first reaction. In the second round of the PCR the amount of the target DNA was increased by the second set of primers, whilst the number of potentially inhibitory factors was decreased.

Subsequently, in order to determine the sensitivity of Chelex/ITS PCR, known trypanosome DNA at different concentrations was spotted onto a FTA card and five punches were selected to be eluted by Chelex; then ITS-nested PCR was carried on the eluted DNA. As can be seen from (Figure 3-6) DNA was successfully amplified in 67% of the twelve dilutions. Positive DNA amplification was detected at a concentration of ~1ng/µl which is estimated as being equivalent to a single trypanosome (Cox, 2007).

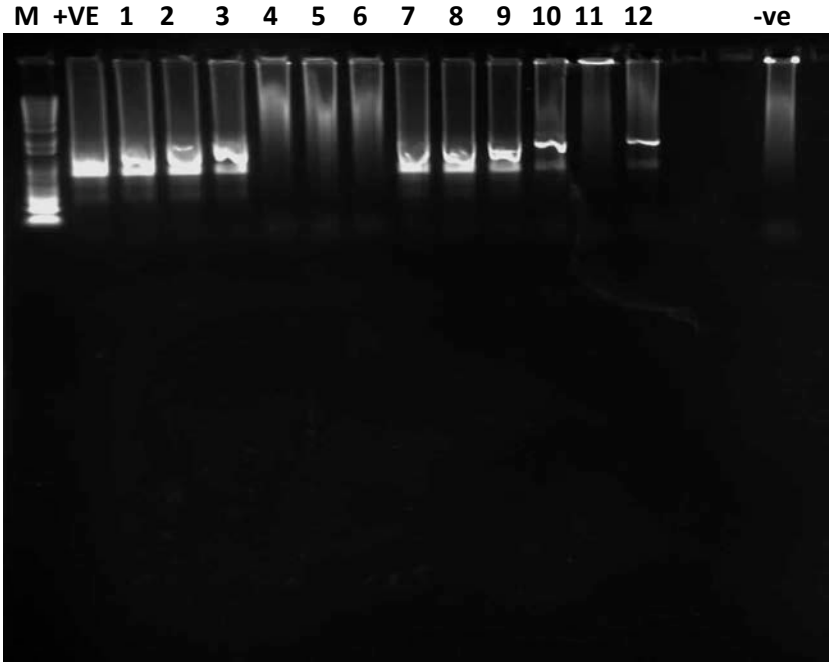


Figure 3-6 Nested ITS-PCR of trypanosome DNA spotted on FTA cards at different concentrations and then extracted using the Chelex protocol. (M) represents the 1kb plus DNA marker, (+Ve) represents *T. brucei* DNA and lanes 1-12 represent trypanosome genomic DNA at different dilutions rom 10^{-1} to 10^{-12} respectively .

Moreover, the relationship between the concentration of the DNA before amplification and the detection of the trypanosome DNA after amplification is shown in (3-7). Which shows that nested ITS-PCR was able to detect trypanosome at 1ng concentration of DNA.

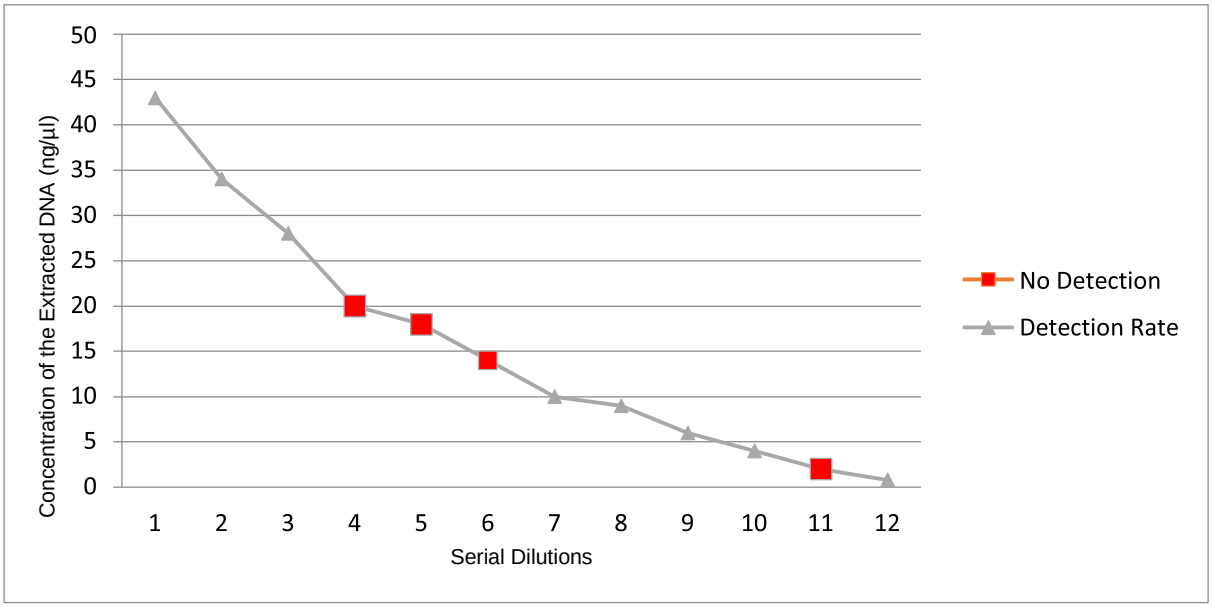


Figure 3-7. The relationship between the concentration of the DNA and generation of the trypanosome nested ITS-PCRproduct.

To test the method on the Angolan samples a few pre-tests need to be carried out. Firstly, it is vital to confirm that the elution of the DNA from the Chelex samples produces DNA that is suitable for amplification using the nested-ITS PCR. Therefore, mammalian tubulin PCR (detects the tubulin gene from the human DNA in the sample) was carried out in the interest of confirming the successful elution of DNA by the Chelex from all Angolan samples. This protocol is often used to verify the integrity of DNA samples. After amplification, it can be seen from (figure 3-8 and 3-9) that 24 out of 36 samples have shown bands of DNA. (Table 3-2) shows the samples that have been successfully amplified. Thus, it can be concluded that a high proportion of the samples are suitable for PCR.

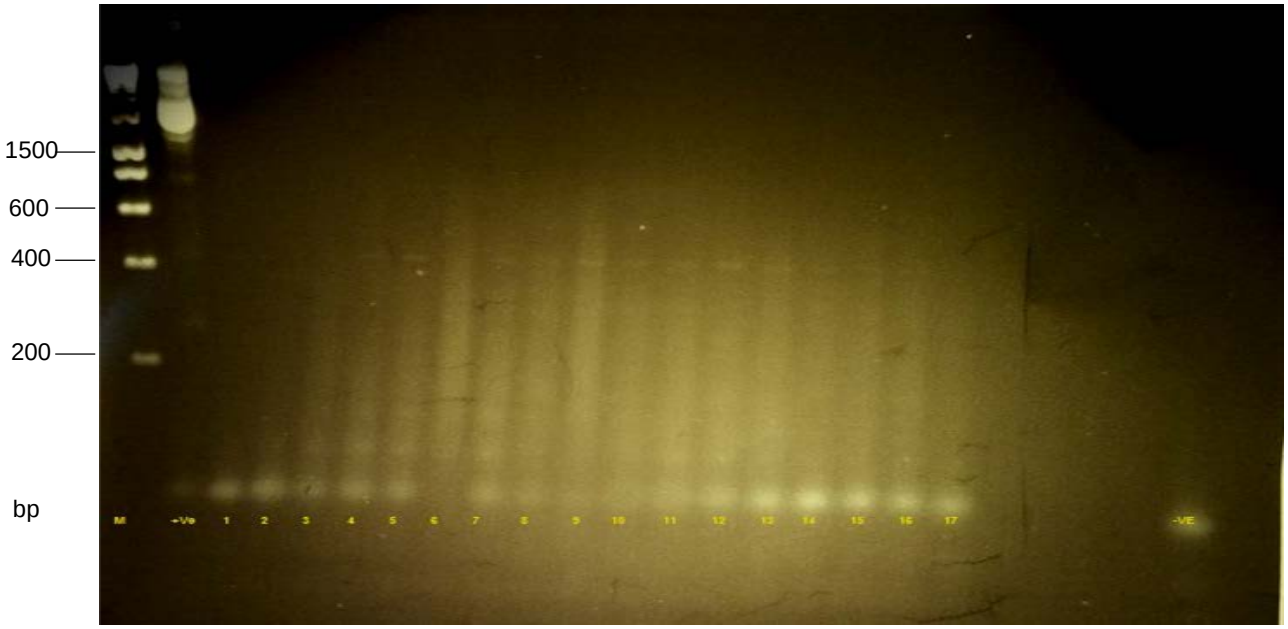


Figure 3-8. Mammalian tubulin PCR amplification of the Angolan samples. Positive DNA control (mouse DNA) is indicated by (+Ve), lanes from 1 to 17 represent the number of the Angolan samples, negative control (water) is indicated by (-Ve) and M represents the 1 Kb (Bioline) DNA marker.

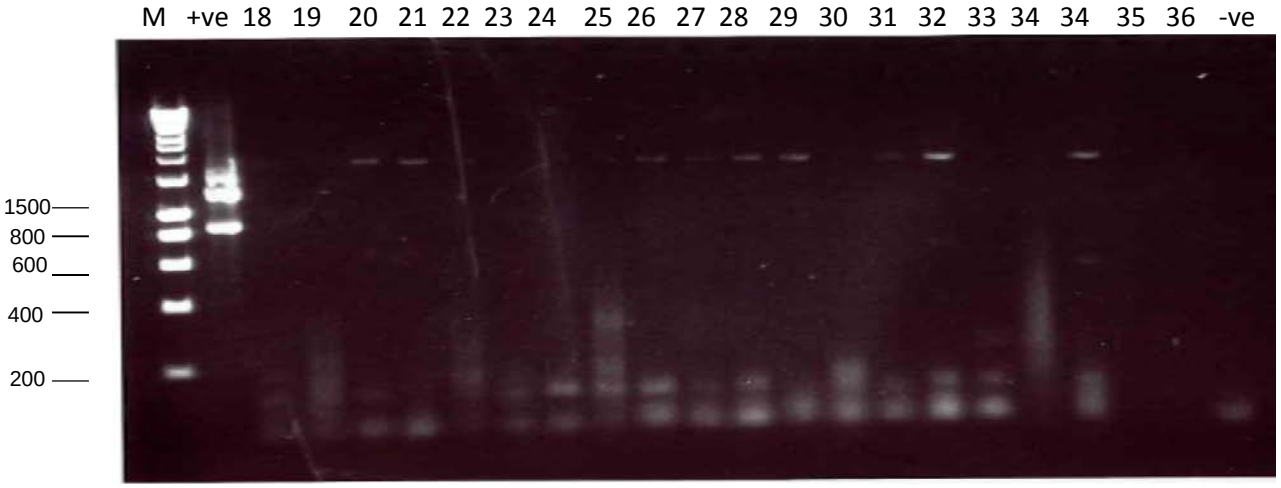


Figure 3-9. Mammalian tubulin PCR amplification of the Angolan samples. DNA positive control (mouse DNA) is represented by +Ve, lanes from 18 to 36 represent the number of the Angolan samples, negative control (water) is indicated by (-Ve), and M represents the 1 Kb (Bioline) DNA marker. The bands show the successful Chelex DNA elution from the punches of the FTA cards.

Sample numbers	Tubulin Positive	Tubulin Negative	Total
1,2,3,6,17,18,19,23,30,33,35,36		√	12
4,5,7,8,9,10,11,12,13,14,15,16, 20,21,22,24,25,26,27,28,29,31,32,34	√		24

Table 3-2. Interpretation of mammalian tubulin PCR amplification

Moreover, all the samples that were collected from Angola were positively tested for trypanosomiasis by CATT. Following the confirmation of DNA elution by Chelex using mammalian tubulin PCR. Although, it confirms that out of the 36 samples only 24 have shown bands in the gel electrophoresis image following the tubulin amplification (Table 3-2). However, the nested ITS-PCR was carried on the all 36 samples. Out of the 36 samples only 13 were successfully amplified, and 63.8% of the samples have failed to amplify. Of the 13 amplified samples, three samples (1, 3, and 17) were recorded as negatives in mammalian tubulin PCR (Table 3-2). The sensitivity of the nested ITS-PCR is measured in accordance with the prevalence of detection of trypanosomes and the positivity of the tubulin PCR. (Table 3-4 below).

Where the prevalence of detection of trypanosomes by nested ITS-PCR was recorded as 11 out of 24 samples, as the samples that have failed to amplify by mammalian tubulin PCR and nested ITS-PCR were excluded. As the failure of the nested ITS-PCR could be due to the absence of template DNA from the samples, which may have caused false negative (figures 3-10 and 3-11). Therefore, the sensitivity of the nested ITS-PCR is measured as (45.8%).

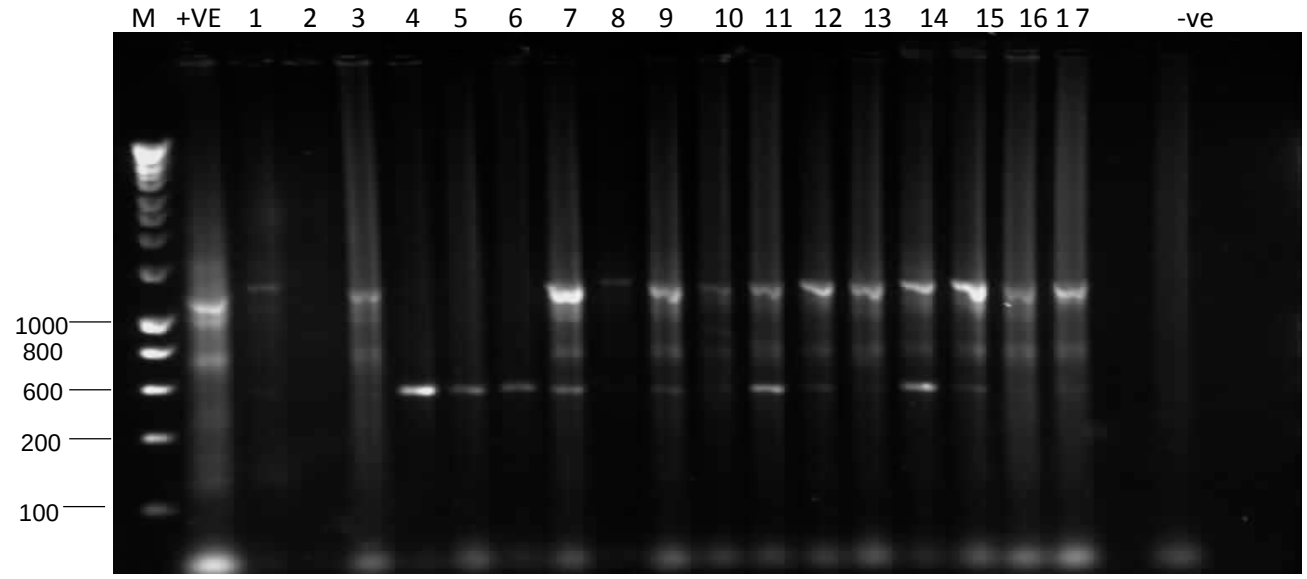


Figure 3-10. Nested ITS-PCR amplification of DNA that was extracted from 5 punches from the FTA cards of the Angolan patients using Chelex® 100. (+ve) represents the positive control of the PCR amplification (*T.b.rhodesienserhodesiense DNA),* lanes 1-17 represent the number of the patient sample, (-ve) represents the negative control (water) and M represents the DNA marker hyperladder 1 (Bioline). Note, samples 4, 5, and 6 amplified an unexpected size of PCR product.

Moreover, considering that only five punches were selected from the FTA card for the Chelex DNA elution. The efficiency of the Chelex DNA elution can be noted from (figures 3-10), as the majority of the DNA amplification of the eluted DNA can be considered as highly concentrated with an exception of sample 1 and 10.

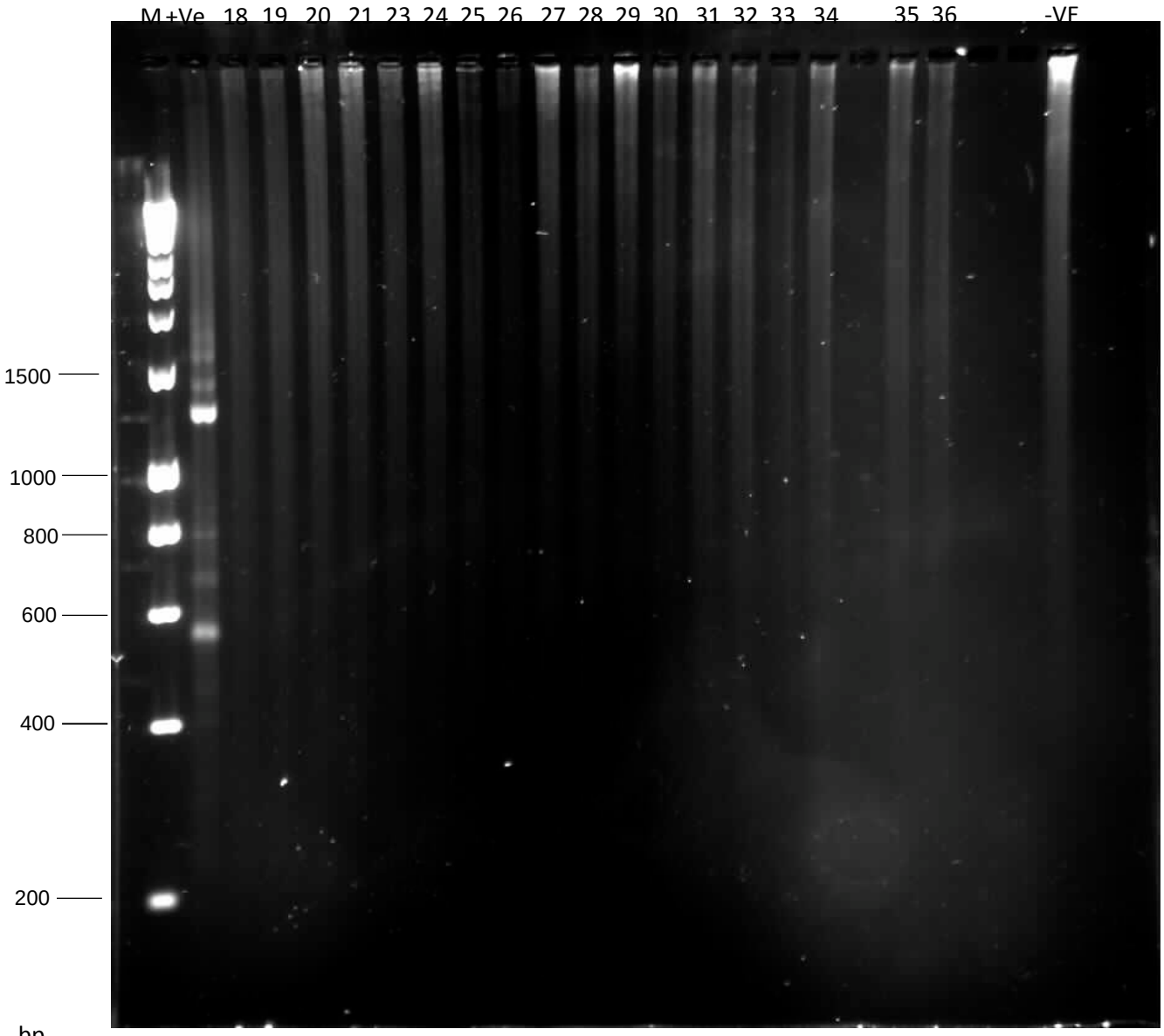


Figure 3-11. Nested ITS-PCR amplification of DNA that was extracted from 5 punches from the FTA cards of the Angolan patients using Chelex® 100 method. (+ve) represent the positive control of the PCR amplification(*T.b. rhodiense* DNA) , lanes 18-36 represent the number of the Angolan patient sample, (-ve) represents the PCR negative control (water).and M represents the DNA marker hyperladder 1 (Bioline)

Although, all the cards were prepared in the same way and collected from confirmed sleeping sickness patients, according to (table 3-2) 12 samples out of the samples in (figure3-11) have been positively tested for DNA presence using mammalian tubulin DNA and the successful PCR reaction as the positive control DNA was successfully amplified after the second reaction. Yet the ITS primers failed to amplify all of the Angolan samples in (figure3-11).

However, at this stage a final judgment on whether it is real negative or false negative cannot be decided. Thus, further investigation is required. Furthermore, the nested ITS-PCR was again carried out on the 17 negative samples (figure 3-11) but this time the number of punches was increased from five to ten punches. This was done to confirm the negativity or positivity of these previously negative samples. However, as shown in (figure 3-12) the post amplification gel electrophoresis image still shows no sign of amplification despite the positive control amplifying correctly.

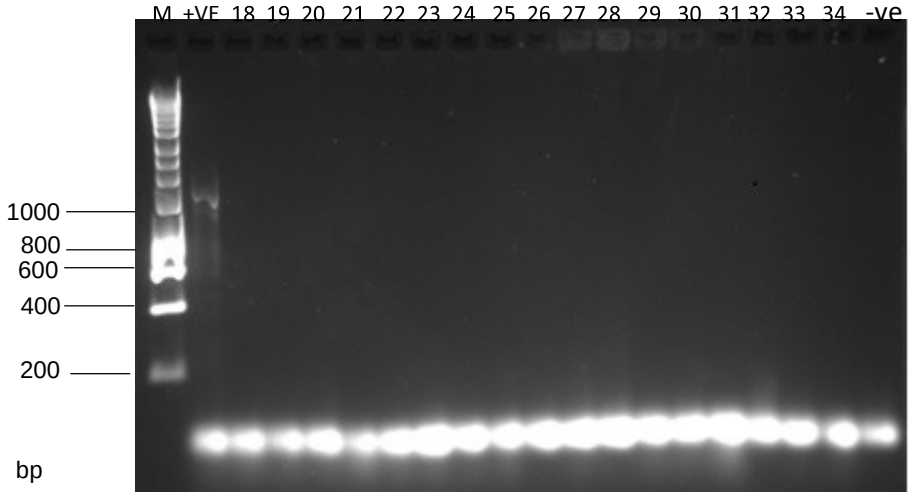


Figure 3-12. Nested ITS-PCR amplification of DNA that was extracted from 10 punches from the FTA cards of the Angolan patients using the Chelex® 100 method from the 18 negative patient samples (Figure 3-11). Marker is represented by (M), +Ve represents the *T.b rhodiense* DNA that was used as the positive control, lanes from 18-35 represent the number of the Angolan sample and –ve represent the negative control (water). Note, although the number of punches has been increased from 5 to 10 in this nested ITS-PCR amplification there remains no trypanosome PCR product.

Furthermore, a different approach was followed by increasing the annealing temperature to 57 Cº. As the GC-rich PCR products are difficult to amplify, increasing the annealing temperature will increase the specificity of the primers. This procedure also failed to produce any amplification (data not shown).

On the other hand, the results of the PCR amplification using ITS PCR of the extracted DNA showed specific size bands for *T.b. brucei*, which was in complete compliance with the expected band size from NCBI database (Table 3-3). However, other unusual sized bands of 600 base pairs were seen in some samples.

Sample Number	Band Size (bp)	Expected Band size/ Species
1	1230	1224/ *T. brucei*
2		
3	1230	1224/ *T. brucei*
4	600	1224/ *T. brucei*
5	600	1224/ *T. brucei*
6	600	1224/ *T. brucei*
7	1230	1224/ *T. brucei*
8	1230	1224/ *T. brucei*
9	1230	1224/ *T. brucei*
10	1230	1224/ *T. brucei*
11	1230	1224/ *T. brucei*
12	1230	1224/ *T. brucei*
13	1230	1224/ *T. brucei*
14	1230	1224/ *T. brucei*
15	1230	1224/ *T. brucei*
16	1230	1224/ *T. brucei*
17	1230	1224/ *T. brucei*

Table 3-3. Expected and observed band sizes resulting from nested ITS-PCR from Angolan patient samples 1-17. The expected sizes were obtained from NCBI database (Cox *et al*., 2005). Patient samples 18 to 36 failed to produce any amplicons (Figure 3-11).

According to (table 3-3) the amplicon sizes resulted from ITS primers amplification of the samples are in complete agreement with the expected band sizes, with exception of samples 4, 5, and 6. Hence, the one of them (sample 4) was sequenced to be identified (below). The prevalence of the detection rate using Chelex/ITS nested PCR and the summary of the results in relation to the patient's occupation, treatment, and staging of the trypanosomiaisis information are illustrated in (Table 3-4).

The staging of the disease is measured based on the variation of the symptoms duration information (Figure 3-4). As Checchi *et al*. (2008b) have reported that duration of stage 1 of Gambian trypanosomiasis is 1.5 years, whereas stage 2 is reported as 3 years. It is worth noticing that patient 8 has 7 years of symptoms duration. Laveran and Mesril (1904) have reported that HAT symptoms can rise to up to seven years. Whereas more recent Sudarshi *et al*. (2014) have reported the longest duration of gambian form of sleeping sickness symptoms (29 years).

Patient Number	Tubulin PCR	ITS PCR	Positive Detection of trypanosomes by ITS-PCR	Occupation	Treatment	Staging
1	-	+	+	Farmer	No	2
2	-	-	N/A	Farmer	Yes	2
3	-	+	+	Farmer	No	2
4	+	+	-	Farmer	Yes	2
5	+	+	-		Yes	2
6	-	+	-	Farmer	Yes	1
7	+	+	+	Farmer	Yes	2
8	+	+	+	Student	Yes	2
9	+	+	+	House wife	No	2
10	+	+	+	Student	No	1
11	+	+	+	Student	Yes	2
12	+	+	+	Farmer	Yes	2
13	+	+	+	Farmer	No	2
14	+	+	+	House wife	Yes	2
15	+	+	+	Farmer	Yes	1
16	+	+	+	Student	Yes	2
17	-	+	+	Farmer	No	1
18	-	-	N/A	Farmer	Yes	2
19	-	-	N/A	House wife	Yes	2
20	+	-	-	Student	Yes	2
21	+	-	-	Farmer	Yes	2
22	+	-	-	Soldier	Yes	1
23	-	-	N/A		No	2
24	+	-	-	Teacher	Yes	1
25	+	-	-	Soldier	Yes	1
26	+	-	-	Farmer	Yes	1
27	+	-	-	Student	Yes	2
28	+	-	-	Soldier	Yes	1
29	+	-	-	Soldier	Yes	1
30	-	-	N/A	Soldier	Yes	2
31	+	-	-	Farmer	No	1
32	+	-	-	Traveller	No	2
33	-	-	N/A		No	1
34	+	-	-	Student	No	1
35	-	-	N/A	Farmer	No	1
36	-	-	N/A	Student	No	2

Table 3-4. Summary of the detection of trypanosomes by Chelex/ITS nested PCR on Angolan patient samples. Patient occupation, treatment history and staging of the trypanosomiasis is specified (1 = early stage; 2 = late stage). (+) and (-) indicate successful and unsuccessful PCR detection respectively, (N/A) not applicable.

According to (table 3-4) overall the Chelex/ITS nested PCR have shown the potentials in term of the sensitivity in detecting the trypanosomes, considering the age of the samples, number of punches and the costs. Further improvements are required to improve the sensitivity.

All the samples that were positively detected by Chelex/ITS-PCR (13 samples) belong to female patients. This could be due to the high intensity of the

parasitemia, because of their occupation as seven out of the 13 (females) works as a farmers, which means high exposure rate to the parasite. Whereas, 3 males in this study were farmers. Also, we cannot be sure whether if the patients have received a treatment or not, have an impact on the detection rate by Chelex/ITS-PCR. Since 7 out of the 13 samples who were positively tested by nested ITS-PCR have received a treatment. Finally, the not applicable is referred to samples who have failed to be amplified in both detection protocols (i.e. tubulin or nested ITS-PCR). These samples will be excluded from the statistical analysis (below).

3.3.3 Identification of nested ITS-PCR amplified products (DNA sequencing)

The nested ITS-PCR amplification produced band sizes of the correct size for *T. brucei* in the majority of positive cases; however three samples (4, 5 and 6) produced a different band size (of 600bp). To find out what these bands were, sample 4 as well as samples 1 and 14 were purified from the agarose gel and sequenced using the internal primers (ITS3 and ITS4). The obtained DNA sequencing results were visualised using the software program Finch TV. Following the viewing of the nucleotides resulting from the DNA sequencing, the sequences were compared to the sequences available on the NCBI database, and a BLAST search was conducted using the National Centre of Biotechnology Information (NCBI) website facility (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?).

The results of the BLAST search showed that sample 1 (which had the *T. brucei* sized band of 1224bp) shared 98% nucleotide similarity with the *Trypanosoma brucei* gene for 5.8S ribosomal RNA (Accession number X02483.1) (Figure 3-13). Sample 14 was in the second band (600bp band) and shared 92% nucleotide similarity with *Trypanosoma evansi* genes for 18S ribosomal RNA gene (accession

number AB551922.1) (Figure 3-14). This confirmed that both of these sequences were derived from the correctly amplified DNA. (Table 3-5) summarises the closest matches for each sequence. Sample 4 (at the unusual 600bp band) detected *Malassezia* genomic DNA containing the 18S rRNA gene (accession number JX993625.1) when exposed to a BLAST search (Figure 3-15).

```
X02483.1          TGATATCCATTATACAAAAAAGAGACTATTTATGTGCATGTATAATTGCACAGTATGCAA 60

  Sample1         TGATATCCATTATACAACAAAGAGCATATTTATGTGCATGTATAATTGCACAGTATGCAA 60
                  ***************** ******  **********************************

  X02483.1        CCAAAAATATACATATATGTTTTACATGTATGTGTTTCTATATGCCGTTTGACATGGGAG 120
  Sample1         CCAAAAATATACATATATGTTTTACATGTATGTGTTTCTATATGCCGTTTGACATGGGAG 120
                  ************************************************************

  X02483.1        ATGAGGGATGTTATATATAGTTCTGTTATTTTCTAATATGTATGTGTGTTAGAGTGTCTG 180
  Sample1         ATGAGGGATGTTATATATAGTTCTGTTATTTTCTAATATGTATGTGTGTTAAAGTGTCTG 180
                  ************************************************** ********

  X02483.1        TGTTAATATACTTTTTAATGCGTGCTCTACATAATATACAGTAGTAATAACACAGAGAAT 240
  Sample1         TGTTAATATACCTTTTAATGCGTGCTCTACATAATATACAGTAGTAATAACACAGAGAAT 240
                  *********** ************************************************

  X02483.1        ACGTATGTAATGCGTATCTCTCTATATCTATATATATATGTATATATGCTATGTGTATAT 300
  Sample1         ACGTATGTAATGCGTATCTCTCTATATCTATATATATATGTATATATGCTATGTGTATAT 300
                  ************************************************************

  X02483.1        AAACCTCGCATATTTTCTCCCTGTTGACCACGGCTCCCACAACGTGTCGCGATGGATGAC 360
  Sample1         AAACCTCGCATATCTTCTCCCTGTTGACCACGGCTCCCACAACGTGTCGCGATGGATGAC 360
                  ************* ********************************************

  X02483.1        TTGGCTTCCTATTTCGTTGAAGAACGCAGCAAAGTGCGATAAGTGGTATCAATTGCAGAA 420
  Sample1         TTGGCTTCCTATTTCGTTGAAGAACGCANCAAAGTGCGATAAGTGGTATCAATTGCAGAA 420
                  **************************** ******************************

  X02483.1        TCATT-CATTGCCCAATCTTTGAACGCAAACGGCGCATGCCCGACGCTCTCTCGAGC-AT 478
  Sample1         TCATTTCATTGCCCAATCTTTGAACGCAAACGGCGCATGGGAGAAGCTCTCTCGAGCCAT 480
                  ***** ********************************   ** ************ **

  X02483.1        CCCCGTACATACCACATTTCTCAGTGTCGAATATAAAAACAAAACACACACCTATTTTTT 538
  Sample1         CCCCGTGCATGCCACATTTCTCAGTGTCGAATATAAAAACAAAACACACACCTATTTTTT 540
                  ****** *** ************************************************

  X02483.1        GTGTTGTTCAACGCACGCACAAAATTCCGCCACCTCTTCTCCTCGTGTGGTGCATATTCA 598
  Sample1         GTGTTGTTCAACGCACGCACAAAATTCCGCCACCTCTTCTCCTCGTGTGGTGCATATTCA 600
                  ************************************************************

  X02483.1        TGTTTGTGAGTGTGCACATATACGATATCATTCTTCTCTTTCTACACGCACGATTGTATA 658
  Sample1         TGTTTGTGAGTGTGCACATATACGATATCATTCAACTCTTTCTACTCGCACGATTGTATA 660
                  ********************************  ********** **************

  X02483.1        TGTCACGCATGTACGTGTGTGTAGTGAGTGATATGGAAGAGAAATGGGAAAGGCATATAT 718
  Sample1         TGTCACGCATGTACGTGTGTGTAGTGAGTGATATGGAAGAGAAATGGGAAAGGCATATAT 720
                  ************************************************************

  X02483.1        ATGTATATGTATATACGTGAT 739
  Sample1         ATATATATGTATATACGTGAT 741
                  ** ******************
```

Figure 3-13. DNA sequence alignment between X02483.1 (*Trypanosoma brucei* 5.8S ribosomal RNA gene) and sample 1. The nucleotide substitutions are highlighted in red and deletions are indicated by (-).

```
AB551922.1      AAACACACACCTATTTT--GTGTTGTTCAACGCACGCAAAAAATCCCGCCACCTCTTTCT 58
Sample14        AAACACACGCGTATTTTTTGTGTTTTACACCGCCCGCAAAAAATCGCGCCACCTCTT-CT 59
                ******** * ******  ***** * ** *** *********** *********** **

AB551922.1      CCTCGTGTGGTGCATATTCATGTTTGTGAGTGTGCACATATACGATATCTTTCAACTCTT 118
Sample14        CCTTGTGGGGCGCATACTCATGTGTGAGAGTGCGCACATATACGATATTTTTCCTCTCTT 119
                *** *** ** ***** ****** ** ***** ************** ****  *****

AB551922.1      TCTACTCGCACGATGGTGTATGTCACGCATATACGTGTGTGTAGTGAGTGATATGGAAGA 178
Sample14        TATACTCCCACGGGTGTATGTCTCACACATACGCGTGTGTGTAGTGAGAGATAGAGAAGA 179
                * ***** ****   ** * * **** ****  ************** ****  *****

AB551922.1      GAAATGGGAAAGGCATATATATATATGTATATACATAATATATATGTGTGTGGATTTGTG 238
Sample14        GAAAGGGGAAAGGCATATATANATAT--ATATACATAATATATATGTGTGTGGATTTGTG 237
                **** ***************** ****  ******************************

AB551922.1      TGTTGAGCACATATAAGGAAAAAGGTTGTGTGTATATACAGAGAGTCTGTGGCGGTTGGG 298
Sample14        TGTTGAGCACACATAAGGAAAAAAATGGTGTGTATATACAGAGAGTCTGTGGCGGTTGGG 297
                *********** **********  * ********************************

AB551922.1      ACATGTGTATAAATATATATATGTATATGTGTGTGTTCCGCTGTGGAGATTTTATATCTT 358
Sample14        ACACGTGTATAAAAATATA--TGTATATGTGTGTGTTCCCCNGTGGAGATTTTATATCTT 355
                *** ********* *****  ***************** * ******************

AB551922.1      ACGGAGAGTGTTCATATATATATGTTTGTACGCATGTATTTTGGCGCCCCGTATAGAGAT 418
Sample14        TCGGAGAGTGTTCATATATATATGTTTGTAGGCATGTATTTTGGCGCCCCGTATGGAGAT 415
                 **************************** ********************** *****

AB551922.1      TAAAAAAGAAGAGAAAAAGTATGCAAAAGAGGCGGCGGATAGTGTGTATGTGTGTATTCA 478
Sample14        TAAAAAAGAAGAGAAAAAGTATGCAAAAGAGGCGGCGGATAGTGTGTATGTGTGTATTCA 475
                ************************************************************

AB551922.1      CAGCAAGCAACTATATTTTGCTGCTTGTGAGTATATGCATATATGTACATTATGTGCTTG 538
Sample14        CAGCAAGCAACTATGTTTTGCTGCTTGTGAGTATATGCATATATGTACATTATGTGCTTG 535
                ************** ****************************************** 
AB551922.1      TGCTTCTTTCGTGTACGCTTCACTTTTTTATATTGCATTTTTCAGACCTGAG--- 590
Sample14        TGCTTCTTTCGTGTACGCTTCACTTTTTTATATTGCATTTT-CAGACCTGAGGGG 589
                ****************************************** **********
```

Figure 3-14. DNA sequence alignment between AB551922.1 (*Trypanosoma evansi evansi* 18S ribosomal RNA gene) and sample 14. The nucleotide substitutions are highlighted in red and deletions are indicated by (-).

```
JX993625.1      ----------------------GTGTCCCTGGCCGCCTACACCCACTATACATCCACAAA 38
Sample4         CANGCCCTACCGAACCCCATCCGTGTCCCTGGTCGCCTACACAAACTATTGATCCG-GAA 59
                                      ********** *********  *****  ****   **

JX993625.1      CCCGTGTGCACTGTCTTGGAGAAAGGCTTCAGAGAAGTTTTTTGTGGCCTCTCTTGGGGT 98
Sample4         CACGTGTGCACTGTCTTGGAGAATGGCTTCCGAGAAGCCATTTGTGGCCTCGCTTGTTGT 119
                * ********************** ****** ******   *********** ****  **

JX993625.1      CTTTCTTCGCTACAAACTCGAATGGTTAGTATGAACGTGGAACTTGGTTGGACCGTCACT 158
Sample4         CTTTCTTCATTACTCACTCAATTGGGCAATATGAATTACCTACTTGGTTGGACCGCCACT 179
                ********  ***  **** * ***  * ******      ************** ****

JX993625.1      GGCCAACAAACTATACACAACTTTCGACAACGGATCTCTTGGTTCTCCCATCGATGAAAA 218
Sample4         GGCCAACAAACTATAACCAACTTATCNNAACTGATCTCTTGGTTCTCCCATGTTGGAAAA 239
                ***************  ******     *** ******************    *****

JX993625.1      ACGCAGCGAAACGCGATAGGTAATGTGAATTGCAGAATTCCGTGAATCATCGAATCTTTG 278
Sample4         ACACAGCCCAACCCGATTCCACGTGTGAATTGCAAAATTCCGTGAATCATCAAATCTTTG 299
                ** ****  *** ****      *********** *************** ********

JX993625.1      AACGCACCTTGCGCTCTATGGCAATCCGTAGAGCATGCCTGTTTGAGTGCCATGAAATCT 338
Sample4         AACGCACCTTGCACTCTATGGTTCTCCGAAGACTTTGCCTGTTACAGTGCCCTGAAATCT 359
                ************ ********   **** ***   ********  ****** ********

JX993625.1      CCCACCCCAAGCGGTTTTTAAATGAAACGGCTTGGCGGATGGGGTCTGGATGGGTGCCTC 398
Sample4         CCCACCCCATACGGGTTTTAAATGAAACGGCTTGCTTGATGGGGTCTGGATGGGCGCCTC 419
                *********  *** ******************   ***************** *****

JX993625.1      TGCCTGCGCTACCTAGCACAGGCTCGCCCGAAATGCATGAGCGCCTTG------------ 446
Sample4         TGCCTGCCCTACCTAACACATGCTCACCCTAAATGCATGAAACCCTTGTTACAANTTTTG 479
                ******* ******* **** **** *** **********   *****

JX993625.1      ------
Sample4         CTTCAA 485
```

Figure 3-15. DNA sequence alignment between sample 4 and JX993625.1(*Malassezia restricta* strain 99B internal transcribed spacer 1).

Sample Number	Band size (bp)	Sequence Length (bp)	Length of covered reference sequence (bp)	Percent identity to the *T. brucei* reference (%)
1	1230	741	919	98
14	600	589	590	92
4	600	485	446	0

Table 3-5. BLAST search results of nucleotide sequences of the ITS-PCR amplified products from three Angolan samples.

According to table 3-5 it can be noticed the resultant sequence length of sample 1 is shorter than the amplified product size. Which could be due to impurities still exist within the purified sample, or mis-priming.

3.3.4 Analysis of positive and negative infections in the Angolan samples

The agreement between the trypanosome detection between the two diagnostic tools (CATT and ITS-nested PCR) was analysed using KAPPA test using Graphpad Prism software; the Kappa value was used to determine the level of agreement between the diagnostic tests. The explanation of Kappa value is described in (Altman, 1991). The KAPPA value was calculated for samples that proved to be positive in the mammalian tubulin amplification. KAPPA value was calculated by Graph Pad Prism 6 and it was found as K= 0.471 which indicates moderate agreement between CATT and nested ITS-PCR/Chelex.

Furthermore, an interesting outcome of this study is that all the positive tested samples by ITS-PCR were females (13 females). In order to find if there is a significance difference between the detection rate in males and females, the P value was calculated to analyse the detection of trypanosomes between the males and females using fisher exact test (95% confidence intervals). P= 0.0001 (<0.05) which indicates significant difference between the detection of trypanosomes in females versus the males, (figure18).

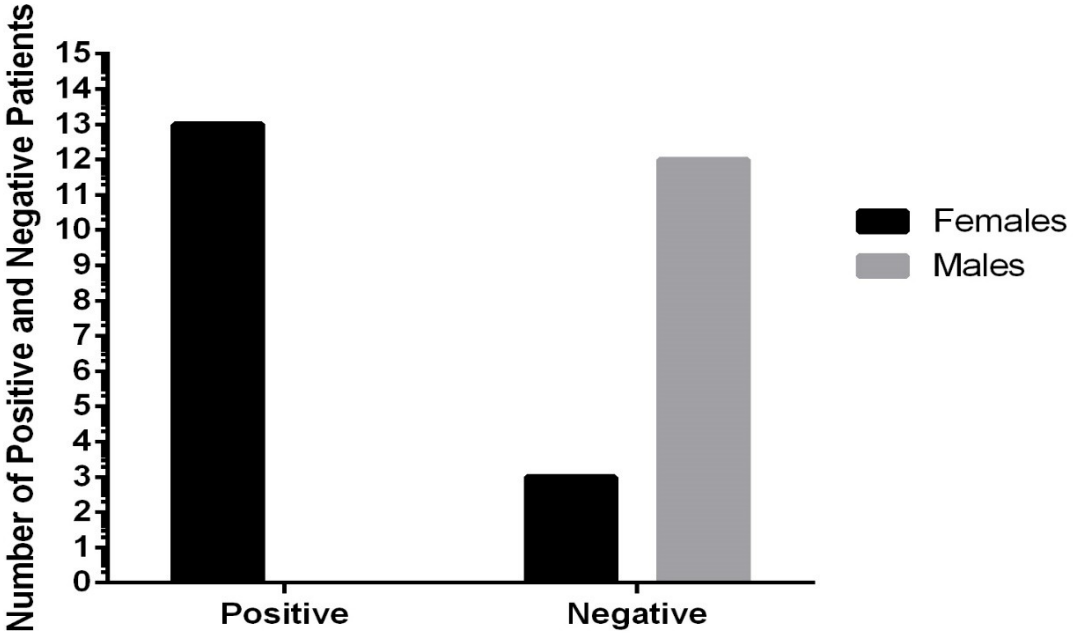


Figure 3-16. The number of trypanosome positives detected by nested ITS-PCR/Chelex between the two genders that participated in this study.

Moreover, according to table 3-4 the majority of the samples that were positively tested by nested ITS-PCR belong to patients that works as a farmers, which agrees with the work that was reported by Truc et al., (2011), where the authors have reported 35% out of the 200 patients infected by gambian HAT were farmers.. Which may have an impact on the detection rate of trypanosomes by nested ITS-PCR. Thus, to test hypothesis of whether the occupation has an impact on the trypanosomes detection rate using nested ITS-PCR. The P value was calculated

using fisher exact test (with 95% confidence intervals) to analyse if there is a significant difference in term of trypanosomes detection between the agricultural and non-agricultural occupations. The P value was calculated as P=0.7051 which indicates there is no significant difference between the mentioned occupations in term of trypanosome detection by ITS-PCR.

A further parameter was investigated which is the relationship between the treatment history (patients who received treatment versus patients who have not received a treatment) and detection of trypanosomes by nested ITS-PCR. (Figure 3-17) shows the detection of trypanosomes by nested ITS-PCR, versus the number of patients who have received a treatment and who have not received a treatment. The P value was calculated using fisher exact test, to determine to see if there is a significance difference between the patients treatment history and trypanosomes detection rate by nested ITS-PCR. The P value was calculated as P=0.228 indicating there is no significant difference between the detection rate of trypanosomes by nested ITS-PCR and patients treatment history.

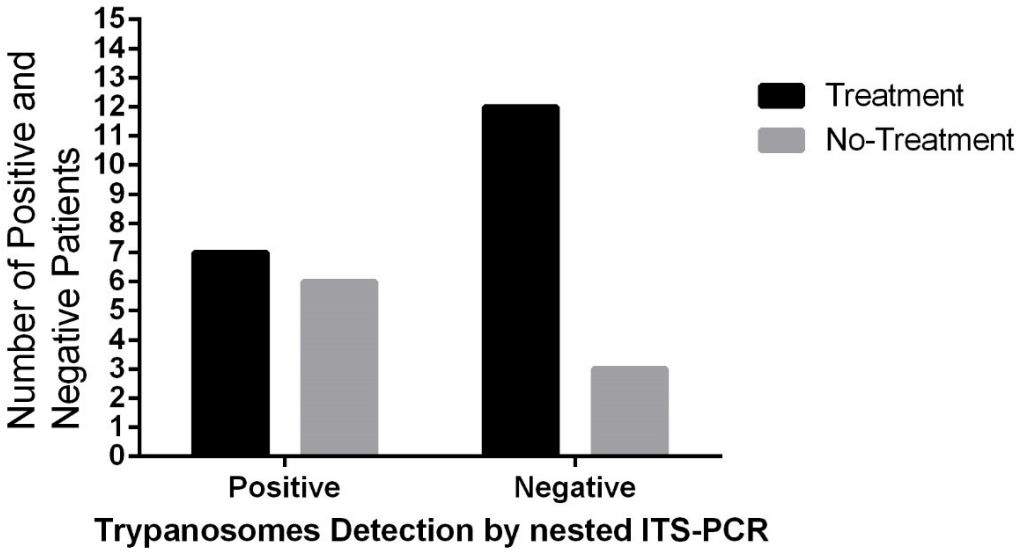


Figure 3-17. The detection of trypanosomes by nested ITS-PCR in relation to patient treatment history.

Finally, the according to (table 3-4) the patients symptoms duration varies, which

reflect the stage of the disease. In interest to study if there is a correlation between the trypanosomes detection rate by nested ITS-PCR and the disease staging, the P value was calculated using fisher exact test (with 95% confidence intervals). The P value was calculated as P=0.434 (P>0.05), which indicate there is no significance difference between the prevalence of detection rate of trypanosomes by nested ITS- PCR and the staging of trypanosomiasis in the studied patients.

3.4 Discussion

Microscopic methods have been used here to detect and differentiate trypanosomes, and these methods have encountered disadvantages in term of sensitivity and time consumption when a large number of samples is involved. This is despite the fact that molecular tools used in diagnosing sleeping sickness have been said by some scholars to be highly sensitive, more specific and less time costly (Desquesnes, 1997; Masake *et al*., 2002). Ahmed (2009) has reported on 84 cattle samples that were tested for any trypanosomes using microscopic methods and PCR; the results were 15.5% and 72.6% detection rates respectively; this constitutes a huge discrepancy in detection.

Previous protocols were only able to target internal ribosomal genes (Mclaughlin *et al*., 1996) through a single round of ITS-PCR on an extracted genomic DNA. Cox *et al*. (2005) have reported that when this protocol was followed on a blood sample spotted onto a filter paper it failed to amplify the DNA. Thus, Cox *et al.* (2005) developed an ITS-nested PCR that was initially found to assist the epidemiological investigations of African trypanosomiasis. ITS-nested PCR has offered a great deal of benefits in terms of cost reduction, and the process is now considered to be a reliable and powerful tool.

The nested ITS-PCR method has not been used on Chelex eluted DNA from FTA cards. The current study was carried out to test the sensitivity and specificity of ITS-nested PCR on trypanosome DNA extracted using the Chelex extraction protocol. The study has thus sought to test the effectiveness and usefulness of this tool using

the Chelex extraction protocol which could be considered to be an alternative and viable method for extracting DNA from FTA cards.

The decision to use a nested PCR was based on the need to increase the sensitivity of the reaction with an increase in the target amount and a potential dilution of the inhibitory factors (Cox, 2007). Initially, to evaluate the nested ITS-PCR, the nested ITS-PCR was carried out on a single punch taken from an FTA card spotted with neat trypanosome DNA, and from FTA cards spotted with blood from patients who have tested positively for trypanosomiasis using CATT (all the sample were from Angola). This resulted in a (0) detection, however, PCR reaction positive control (*T.b.rhodesiense* DNA) have been successfully amplified. This accords with the findings reported by Cox *et al*. (2005) which showed that a single punch from an FTA card can cause false negative PCR results due to the uneven distribution of the genomic material on the FTA card matrices. The sensitivity of the nested ITS-PCR increased suddenly by using the Chelex DNA elution approach. Chelex/nested ITS-PCR was studied on serial dilutions (1:12) of neat trypanosome DNA which was spotted onto FTA card and eluted using the Chelex, whereby it recorded a sensitivity of 67% out of the twelve dilution. Which could be due to the absence of template DNA, which may be due to the failure of Chelex DNA extraction. From Figure 3-7 it can be clearly seen that nested ITS-PCR was highly sensitive and trypanosome was detected at a concentration of 1ng/µl.

Based on the tubulin mammalian PCR results it can be concluded that using Chelex to extract DNA is a reliable and convenient approach. The key advantage of using Chelex®100 is that it is cheaper when extracting DNA, with an estimated cost of less than US$0.16 per sample for the reagents compared to US$1.23 per sample for

InstaGene Matrix (Strøm *et al*., 2014). Moreover, Chelex®100 possesses metal ions which might act as catalysts in the breakdown of the DNA at high temperatures in low ionic conditions, and this can provide protection to DNA from degradation resulting from high temperatures.

Only 24 samples out of the total of thirty six were positively confirmed in terms of the presence of DNA using mammalian tubulin PCR. The Chelex/nested ITS-PCR was carried on the 36 Angolan samples. The samples that tested negative when using mammalian tubulin PCR were excluded from the analysis of the detection rate by nested ITS-PCR, as they could lead to false negatives. This was due to the uncertainty of the presence of the template DNA. Thus, out of the 28 final samples (see Table 4), only 11 (45.8%) samples were successfully amplified, whereas 15 (53.6%) failed to amplify. The results recorded a higher detection rate in contrast with the results that were reported by Thumbi *et al*. (2008) who tested 103 cattle blood samples infected with trypanosomes, using ITS single PCR and nested ITS PCR and species-specific PCR; these detection rate results were 26.2%, 28.1% and 10.7% respectively. The ITS nested PCR method was able to detect 3.8% of the mixed infections compared to 1.9% when using a single ITS PCR. These findings are comparable to the results reported by Ahmed *et al*. (2011) whereby the authors used Chelex elution to elute DNA from 300 cattle blood samples infected by trypanosome, to be tested using TBR-PCR (The TBR universal primers developed to amplify the 177 bp sequences from Trypanozoon genomic DNA). Ahmed *et al* (2011). Reported a 56.4% sensitivity.

The Chelex/ITS-PCR failed to amplify all the samples (as in Figure 3-11), despite 12 out of the 18 samples in (Figure 3-11) having tested positive for DNA presence using

mammalian tubulin DNA. According to Becker *et al*. (2004) a single punch "contain"1.34µl of template DNA. Ahmed *et al*. (2011) have reported an increase in sensitivity from 5.9% to 35.6% resulting from an increase in the number of punches from 5 to 10 punches respectively, using TBR-PCR to detect trypanosomes. Thus, the number of selected punches was increased to 10 punches, in order to ascertain a false or genuine negativity. The negative samples still failed to amplify despite this increase in sample size. Thus, a different approach using an increased annealing temperature (raised to 57C°) was used (data not shown); this also failed to amplify. The negative results can be explained by several different factors; firstly, according to Becker *et al.* (2004), storage time has an impact on the FTA cards, as the DNA faces difficulties when released from the FTA matrix. Also, removing all the sample's impurities was a challenge, because this study's samples were collected in the 2001. Moreover, the results may have been due to the presence of PCR inhibitors in the blood samples such as immunoglobulin G in plasma, haemoglobin in erythrocytes and lactoferrin in leukocytes. Secondly, the results may be due to the contamination of the sample by exogenous DNA or DNases which would provoke a destruction of the DNA (Desquesnes, 1997; Desquesnes and Davila, 2002; Solano *et al*., 2002). Reithinger *et al*. (2000) have detected leishmania species in dog blood using PCR in Chelex®100 extracted samples. The authors reported that washing the FTA cards *before* PCR amplification did *not* remove all the PCR inhibitors, since none of the Chelex®100 extracted samples amplified the target DNA. They attributed this failure to the presence of PCR inhibitors that were not successfully removed using Chelex®100. In addition, Singh Bal *et al*. (2014) have detected *T. evansi* DNA from infected mice blood and purified trypanosomes, the authors have reported a higher sensitivity of using ESAG6/7 primers than nested ITS-PCR. The authors have suggested that this could be due to a higher similarity between the ESAG6/7

primers and the target sequence in the genomic DNA of *T. evansi*. Considering that copy number of ITS regions is (~100) higher than ESAG (20). Furthermore, Truce et al. (2012) have reported that the low parasitemia in the blood and CSF is well known for *T. b. gambiense*. Therefore, the amount of parasite DNA extracted from body fluids should be a key factor in terms of sensitivity of PCR using microsatellites, which has been postulated to be a major factor for the analysis of blood samples from patients (Koffi *et al.,* 2007).

The first demographic parameter was tested is the occupation (agriculture and non-agriculture) of the patients. According to the subsequently calculated P value ($P=0.705>0.05$) the statistical analysis indicated there was no significant difference between samples from agricultural workers and those from non-agricultural workers in trypanosomes detection using nested ITS-PCR. The enormous variation in the detection rate of trypanosomes between male and female patients using nested ITS-PCR may be due to the patients' different stages of trypanosomiasis. Truc *et al*. (1999) have demonstrated the valuable role of PCR in studying the *staging* of Gambian trypanosomiasis when using cerebrospinal fluid (CSF) samples from 20 patients from the Côte d'Ivoire. According to the white blood cell counts, five patients out of the 20 were at an early stage of trypanosomiasis. The authors reported that *Trypanosoma brucei*-specific PCR failed to detect the trypanosomes in the 5 patients who were at the early stages of trypanosomiasis. By staging the disease (stage 1 or

stage 2) in the thirty six patients in accordance with their symptoms duration, the P value was calculated to determine if there is a correlation between the trypanosome detection rate using nested ITS-PCR and the disease stage. The P value was recorded as >0.05, which indicates there is not significant correlation between the staging of the disease and molecular detection of trypanosomes by nested ITS-PCR. Which agrees with a study carried by Deborggraeve *et al*. (2010) where the authors reported non-significant difference between stage 1 and stage 2 of patients infected with *T.b.gambiense* (P= 0.184) which indicates the drawback of PCR in staging of gambien HAT .The final demographic parameter (treatment history) was studied in relation to the trypanosomes detection rate in patients when using nested ITS-PCR. Although the positive detection rate was recorded as relatively equal between patients who have received treatment and those who have not received treatment the P value was calculated as >0.05, indicating that there is not a significance difference between patients who have received treatment and those who have not in terms of trypanosomes detection using nested ITS-PCR.

Many publications e.g. (Daniel and Linardi, 2010; and Morrison *et al*., 2007) have suggested that PCR sensitivity can be increased through Multiple Displacement Amplification (MDA). MDA is a protocol that uses the bacteriophage ø29 DNA polymerase and exonuclease-resistant, thio-phosphate-modified and degenerate hexamers to amplify DNA from crude or pure sources. Daniel and Linardi (2010) compared the PCR amplification of DNA from siphonapterids before and after MDA, and the authors found that out of 26 samples only 15.4% were amplified, whereas after the involvement of MDA the sensitivity had increased to 92.2%. Interestingly one of the amplified samples after the MDA was collected as far back as 1909. Morrison *et al* (2007) have reported on the use of MDA on extracted DNA from FTA

cards spotted with blood infected with trypanosome before PCR amplification. The authors found that by using 5 discs from the FTA card the nested PCR was only able to detect 1 trypanosome per reaction without the use of MDA, whereas when the MDA was applied before the amplification, the sensitivity increased by twenty fold (~0.05 trypanosome per reaction).

Several approaches can be considered for increasing the sensitivity of the negative results of the nested ITS-PCR, such as lysing the whole blood before applying it onto the FTA card. Ahmed (2009) compared the sensitivity of DNA elutes from whole and lysed blood extracts, and the author reported that the PCR on 5 µl *lysed* blood elute was 1.3 times more effective than PCR on 5µl *whole* blood elute. This was said to be due to the fact that the lysing protocol led to an increase of the genomic material, and this resulted in a greater degree of homogeneity in the distribution of trypanosome DNA across the FTA card.

Cox *et al.* (2005) have confirmed the specificity of the ITS primers, but samples 4, 5 and 6 have produced amplicons with unexpected sizes after ITS-PCR amplification (see Figure 10a). A DNA sequencing of sample 4 shows the produced band belongs to the *Malassezia* genomic DNA containing the 18S rRNA gene. *Malassezia* are known as commensal yeasts and are found on the sebaceous areas of human skin. Although very resilient, they are generally found as human skin flora, but these can also be considered as a pathogenic organism which is related to several dermatological conditions with substantial cosmetic consequences (Levin and Delano, 2011). It is presumed that somehow this DNA or these organisms made it onto the FTA cards, perhaps from the sleeping sickness patients themselves. This reflects the evident ability of the ITS primers to detect *non*-trypanosome DNA. DNA sequencing of the selected bands confirmed the detection of *T.b. brucei*. The

nucleotides of the sequenced sample 1 have shown a high percentage of alignment with *Trypanosoma brucei* gene for 5.8S ribosomal RNA. All the positively detected samples when using nested ITS-PCR have produced more than one band, except samples 1 and 8. This reflects the usefulness of nested ITS-PCR in detecting mixed *trypanosoma* species. The second band was produced in sample 14 of 600bp size have been sequenced. Which have shown high similarity with *Trypanosoma evansi* genes for 18S ribosomal RNA gene.

In summary, the objectives of this study was to determine the usefulness of nested ITS-PCR in detecting trypanosomes from FTA cards, to evaluate the use of FTA cards in collecting samples in field diagnostic laboratories, the efficiency of Chelex in overcoming the obstacle of false negative results and the correlation between the demographic information and the detection rate of trypanosome by nested ITS-PCR It is concluded that FTA cards are more convenient for collecting larger samples and nested ITS-PCR is a simple, more economically viable, less time consuming and more sensitive process for detection because it has the ability to detect mixed infections. Furthermore, protocols such as MDA and blood lysing could contribute to an increase in the sensitivity of the nested ITS-PCR. Future work is required to test this provisional hypothesis.

Chapter 4 . Identification of trypanosomes from FTA cards spotted with blood using MGE-PCR

4.1 Introduction

Trypanosomiasis is known to be a complex disease as the human-infective and non-human-infective strains of *Trypanosoma brucei* overlap in the same transmission cycles. The epidemiology of HAT is largely based on the identification and tracking of strains involved in the generation of epidemic and endemic foci (Fevre *et al*., 2001; Welburn and Odiit, 2002). Previous studies have introduced molecular tools for tracking trypanosome strains including the use of isoenzyme characterisation (Gibson *et al*., 1980; Tait, 1980, Tait *et al*., 1983, 1985; Stephens *et al*., 1993) and RFLP analysis (Hide *et al*., 1989, 1991, 1994, 1998, 2000). Although these tools have a successful track record in studying trypanosomiasis, they are not efficient when large-scale uses of samples are involved, as they require a significant amount of time for the growth and amplification of parasites in experimental cultures or animals. Thus, they lack the ability to enable a sensitive discrimination of strain type directly from the infected sample.

Hasan *et al.* (1984) have reported the discovery of mobile genetic elements (MGEs) including the ribosomal inserted mobile element (RIME) in trypanosomes. The mobile genetic element PCR (MGE-PCR) was developed as a tool for strain identification of trypanosomes (Tilley *et al.,* 2003) and for the diagnosis of *Toxoplasma gondii* (Terry *et al*., 2001). MGEs are found present in an extensive variety of living organisms from protozoa to humans (Hartl *et al*., 1997). The importance of MGEs relates to their repetitiveness and variability, which can help in improving the efficiency of molecular epidemiological investigations (Hide and Tait, 2009). Also, the ability of MGEs to move to a new chromosomal location can be deployed to study the shaping of the evolutionary process of the genome

(Bhattacharya *et al*., 2002). Moreover, MGEs can be found in high frequencies (Hartl *et al*., 1997) as about 35% of the human genome consists of MGEs and they occupy 10% of the *Drosophila* genome (Labrador and Corces, 1997). On the other hand, about 5% of the *T. brucei* genome consists of MGEs (INGI/ RIME) (Bhattacharya *et al*., 2002). Therefore, MGEs constitute an attractive marker to be used when studying the variability and characterisation of strains of parasites. Based on the fact that MGEs can be used as genetic markers in epidemiological studies, Hide and Tilley (2001) and Tilley *et al.* (2003) have developed a tool known as MGE-PCR, which is a simple and sensitive technique that can be used to detect genetic variability in *Trypanosoma brucei ssp*. MGE-PCR utilises a single primer that produces variable banding patterns in a one-step PCR approach. It is based on using markers derived from the trypanosomal mobile genetic RIME and detecting positional variation in these elements. The MGE-PCR approach is based on the technique of PCR using a single PCR primer method that was reported by Screaton *et al.* (1993) and which facilitates the amplification of unknown ends of genes from known internal fragments. MGE-PCR is considered considerably more specific than the RAPD method, since it is not affected by contaminating DNA to the same extent (Hide and Tait, 2009). It can be used to detect the positional variation of MGEs by enacting PCR amplification from within a known piece of DNA (the MGE such as RIME) into an unknown flanking sequence. This is because there are specific annealing requirements for the annealing of the single MGE-based primer on at least one end of each PCR-amplified fragment (Hide and Tilley, 2001). Furthermore, Tilley *et al.* (2003) have reported the successful use of MGE-PCR in producing a clear distinction between human-infective and non-human-infective

isolates from *T.brucei* stocks collected from southeast Uganda, Kenya and Zambia in the presence of vertebrate (hosts) DNA. Moreover, Njiru *et al.* (2011) have reported that with MGE-PCR the author was able to distinguish between *T. evansi* and *T. brucei.* In addition, Simo *et al.* (2005) applied MGE-PCR to *T.b.gambiense* isolates from Central Africa and they were able to report micro genetic differences between *T. b. gambiense* isolates. The adaptability of this approach is demonstrated by additional studies that have reported the use of MGE-PCR in studying the genetic variability in *Toxoplasma gondii* (Terry *et al*., 2001) and the *trematode Clonorchis sinensis* (Lai *et al.,* 2008).

Moreover, one of the advantages of MGE-PCR is its reproducibility as Tilley *et al.* (2003) have reported that an identical banding pattern was generated from DNA extracted from procyclic isolates collected from Tororo in 1989 and bloodstream isolates collected in 2000 from the same focus. This is an important part of such studies that involve comparing strains from new and established foci (archived DNA), in order to determine the origin(s) of outbreaks, the spread of diseases and the establishment of new foci. Another technique, RAPD, can produce the same level variation but it can only work on purified parasite DNA preparations. MGE has the advantage of specifically amplifying the trypanosome DNA in the presence of the host DNA (Simo *et al*., 2005). MGE-PCR has thus been widely used as a technique using extracted DNA as the source of DNA. However, many recent studies now use blood samples stored on FTA cards as a method of sampling. There are no reports, as far as we know, of MGE-PCR being successfully used to amplify trypanosome DNA and genotype it directly from FTA cards.

There are certain sampling problems associated with the detection of trypanosomes using FTA cards. As Cox *et al*. (2010) have reported, Multiple PCR samples taken

from single FTA cards have demonstrated that a single punch from an FTA card is not sufficient to confirm the infectivity status of an individual animal, as parasite DNA is unevenly distributed across the card. At low parasite densities in the host, this stochastic sampling effect results in an underestimation of the disease's prevalence based on a single punch PCR test. However, Ahmed *et al*. (2011) have reported an increase in the sensitivity of PCR amplification *after* DNA elution from the FTA card by Chelex. The sensitivity increased from (35.6%) to (56.4%) respectively.

Therefore, in this present study the Chelex elution approach will be carried out on blood spotted on FTA cards of patient samples collected from Angola, prior to the MGE-PCR amplification.

4.2 Objectives

1. To execute a detailed analysis of MGEs in relation to the trypanosome genome and strain types in a collection of samples from Angola.
2. To investigate and develop the use of MGEs for field trypanosomes samples collected on FTA cards
3. To compare the Chelex-nested ITS PCR and the MGE-PCR as methods for trypanosome detection and strain evaluation

4.3 Methods

In this chapter a different trypanosome molecular identification approach (MGE-PCR) will be carried out. Detection of trypanosomes was carried out using MGE-PCR, using a REV B primer to amplify the RIMEs from the eluted DNA from FTA cards. The samples were taken from 36 patients who had tested positively for trypanosomiasis by CATT collected from Angola. Furthermore, the resultant amplicons from the PCR amplification will be sequenced to identify the generated bands. Detailed materials and methods can be found in chapter 2.

4.4 Results

Although all the samples, obtained on FTA cards from a hospital in Angola, were confirmed as positive for trypanosomiasis by CATT (see Chapter 3), there were variable results when the samples were tested using the ITS PCR-based diagnostic tests. Adequate DNA extraction from the FTA card punches was confirmed using tubulin PCR, however, nested ITS-PCR failed to detect 14 of the tubulin positive samples. To further examine this issue, it was therefore decided to test the samples by MGE- PCR, as a way of comparing another molecular detection system to CATT.

4.4.1 Development of the use of MGE-PCR in FTA cards

As far as we know, MGE-PCR has not been successfully used directly from FTA cards. There have been attempts made to get this to work from single punches in the past, but these have been unsuccessful. Based on the problems identified by Cox *et al.* (2010) on sampling trypanosomes from FTA cards and the solutions based on Chelex extraction provided by Ahmed *et al.* (2011), Chelex-based approaches were investigated as a method for improving the use of FTA cards for MGE-PCR. Thus, as a first step it was important to evaluate and optimise the technique, before testing the samples. Initially, the basic protocol (Tilley *et al*., 2003) was followed for the use of MGE-PCR amplification from DNA eluted by Chelex from FTA cards. This involved using eluted trypanosome DNA from the FTA card as a positive control. The REV B primer failed to amplify the samples including the positive control (Figure 4-1). This could have been due to the genomic DNA of the host being present on the spotted FTA card causing the REV B primer to encounter difficulties in annealing with the template. A possible solution, therefore, could be to increase the primer concentration to try to optimise the PCR reactions. The concentration of REV B was

doubled to 20pmol. This increase in the REV B primer concentration helped to amplify the positive control samples only (Figure 4-2).

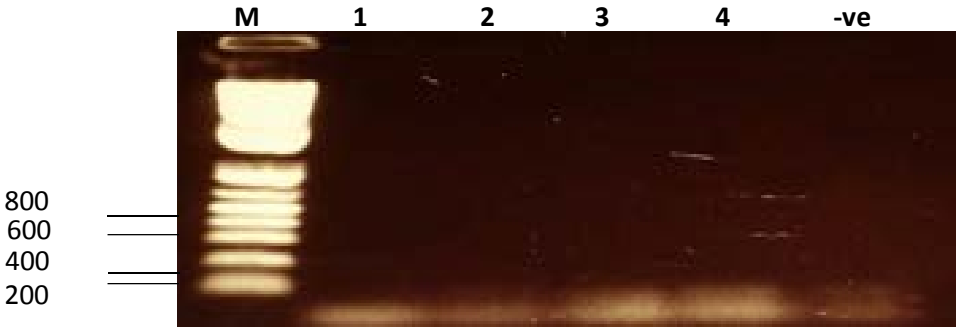


Figure 4-1. MGE-PCR amplication using 10pmol of primer REV B. (M) contains the DNA size marker (1kb), , lane 1 is the positive control (eluted trypanosome DNA from the FTA card), lanes 2-4 (Angolan samples 4, 10 and 20 respectively) and (-ve) is the negative control (water)

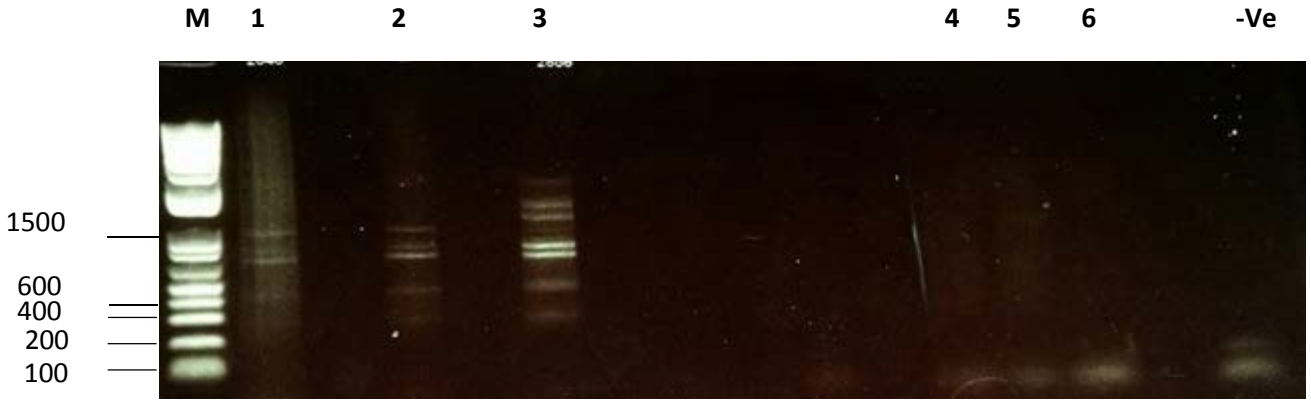


Figure 4-2. MGE-PCR amplification using REV B primer at the increased primer concentration of 20pmol. (M) contains the DNA marker (1kb), lanes labelled 1, 2 and 3 indicate the positive control trypanosome DNA that was eluted from FTA cards by Chelex, lanes 4,5 and 6 represent the Angolan samples (4, 10 and 20 respectively), and (-ve) is the negative control (water).

The results of the MGE-PCR using double primer concentrations (as in Figure 4-2) shows that the amplification when using a single primer works efficiently with an increased concentration but it still fails to amplify from the real samples eluted from the FTA cards. However, the MGE-PCR have successfully worked in the latter experiments with an increased concentration of the primer (below).

4.4.2 Chelex-MGE-PCR detection of trypanosomes in FTA cards

After this initial evaluation of the technique, whereby the concentration of the REV B primer was increased to 20pmol rather than 10pmol, MGE-PCR amplification was carried on eluted DNA from FTA cards collected from sleeping sickness patients in Angola. At a concentration of 20pmol, MGE-PCR using primer REV B resulted in a variation in the banding patterns (Figures 4-3 and 4-4). Of 36 samples, only 21 were successfully amplified using REV B, which was nevertheless higher than the number of the amplified samples using Chelex / ITS nested PCR (Chapter 3). DNA molecular size was measured based on the migrated distance of the band, illustrated by comparing each band against its size in the DNA marker (see Table 4-1). It can be clearly seen from (Figure 4-3 and 4-4) that all the samples share the same banding pattern, which is different to the positive control as it was a *T. b. rhodesiense* stock from Uganda, not a *T. b. gambiense* isolate that would be expected in Angola.

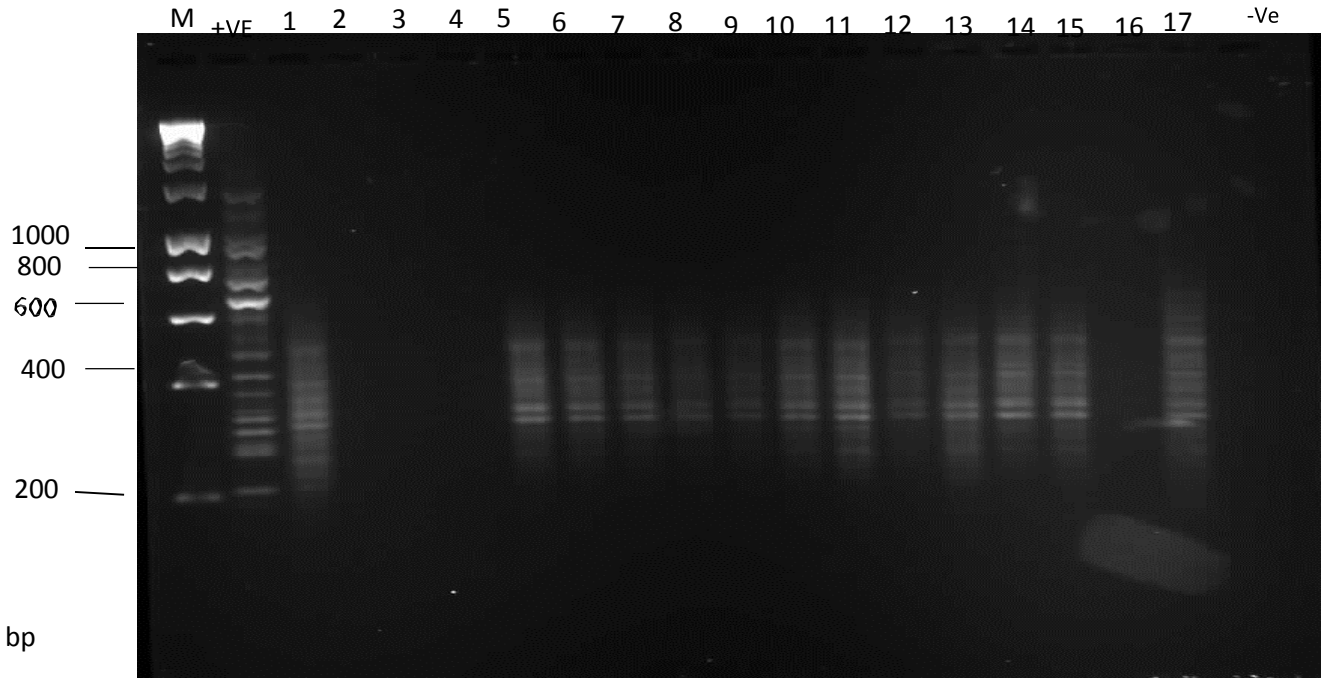


Figure 4-3. MGE-PCR of FTA card samples from Angolan sleeping sickness patients. . (M) represents the Hyperladder 1 size marker (Bioline), (+Ve) represents the positive control (*T.b.rhodiense* DNA), lanes 1-17 represents the number of the patient samples and (-Ve) indicates the PCR negative control (water).

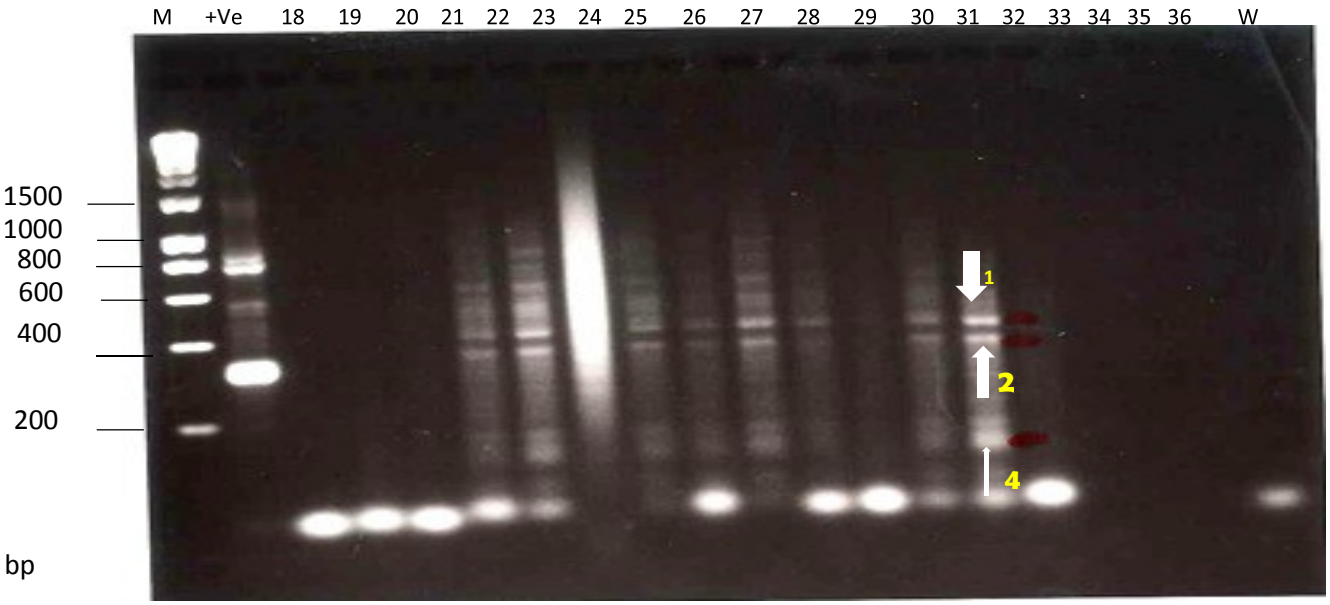


Figure 4-4. MGE-PCR of FTA card samples from Angolan samples sleeping sickness patients. (M) represents the Hyperladder 1 size marker (Bioline), (+Ve) represents the positive control (*T.b.rhodiense* DNA) and lanes 18-36 represent the number of the patient and (-Ve) represents the negative control (water). Bands selected for DNA sequencing are highlighted by the white arrows.

Table 4-1. Summary of the MGE-PCR data from the Angolan samples. The presence and the absence of the bands between 200bp and 2000bp are represented by + and – respectively.

DNA Maker Sizes (bp)	200	220	240	260	280	300	320	340	360	380	400	420	440	460	480	500	520	540	560	580	600	620	640	660	680	700	720	740	760	780	800	820	840
Sample Number																																	
+Ve	+	-	-	+	+	+	-	-	-	+	+	+	-	+	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-
1	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-
2																																	
3																																	
4																																	
5	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

DNA Maker Sizes (bp)	860	880	900	920	940	960	980	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
Sample Number																		
+Ve	-	-	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	+
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2																		
3																		
4																		
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

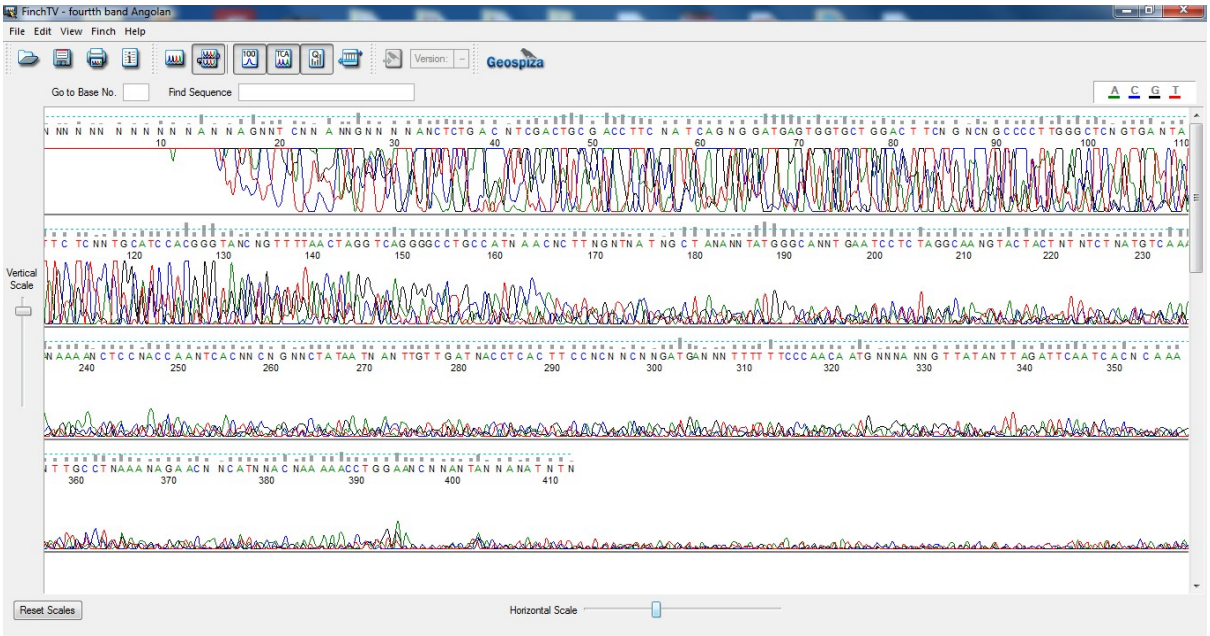
DNA Maker Sizes (bp)	200	220	240	260	280	300	320	340	360	380	400	420	440	460	480	500	520	540	560	580	600	620	640	660	680	700	720	740	760	780	800	820	840
Sample Number																																	
+Ve	+	-	-	+	+	+	-	-	-	+	+	+	-	+	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-
18																																	
19																																	
20																																	
21	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	+	+	-	-	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
35																																	
36																																	

DNA Maker Sizes (bp)	860	880	900	920	940	960	980	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
Sample Number																		
+Ve	-	-	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	+
18																		
19																		
20																		
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

According to Table 1 all the amplified samples share the same pattern. The Angolan samples appeared to have very low molecular weight bands, with 460bp constituting the highest molecular weight, as compared to the *T. brucei* control.

4.4.3 Investigation of REV B specificity using DNA sequencing to confirm band sizes

To our knowledge, this is the first time that RIME positional variation has been demonstrated using a field sample (FTA cards). Thus, it is vital to analyse the amplicons generated by REV B amplification by sequencing so to ensure that they are the desired product. Selected bands from sample 31 (in Figure 4-4 marked with white arrows) were cut and purified to be identified through DNA sequencing (see Figure 4-5).



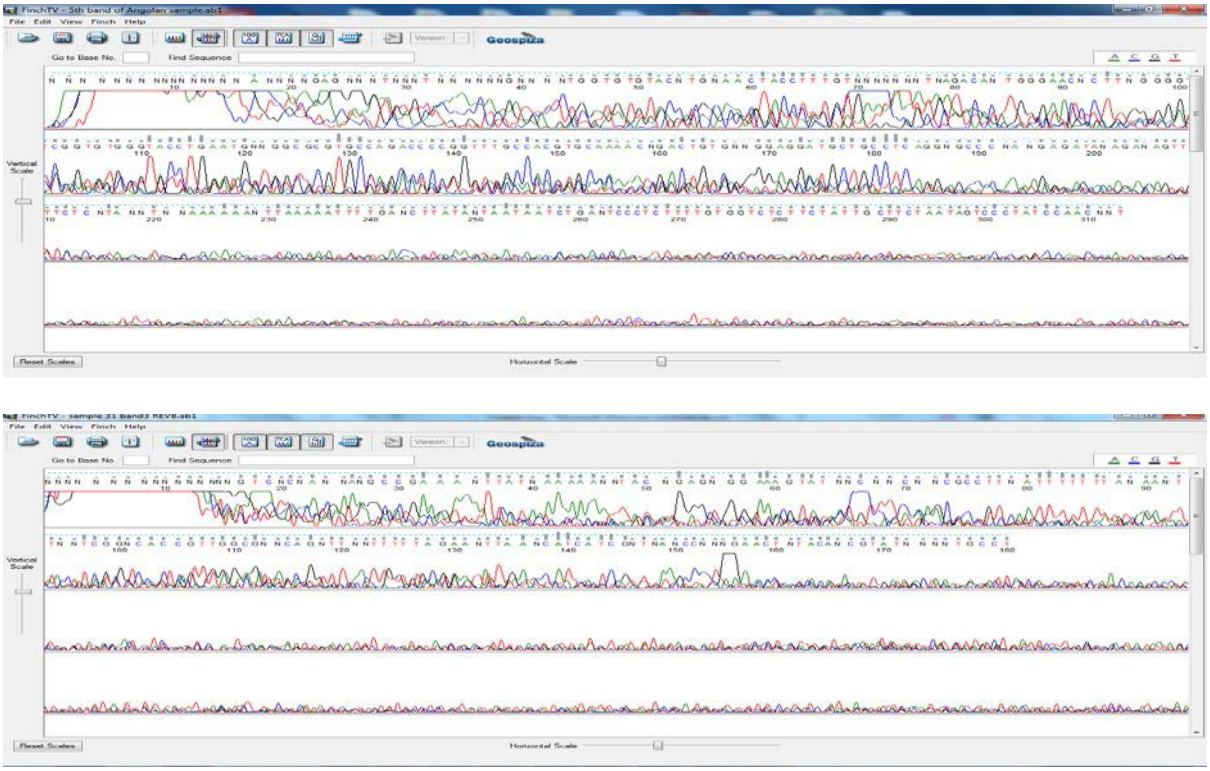


Figure 4-5. Screen shots of the DNA sequencing of REV B amplified products 1, 2 and 4 (see Figure 4-4) derived from Angolan sample 31. The figures are arranged in accordance with the molecular weight of each band (from the highest to the lowest). The DNA sequencing was analysed using FINCH TV software. In all cases, the DNA sequencing was of very low quality.

As can be seen from (Figure 4-5), the DNA sequencing of the MGE-PCR products was not efficient or readable. This could perhaps have been due to REV B mis-priming during the sequencing process, the samples being contaminated, or the primer binding in a non-specific way.

4.4.4 Improving DNA sequencing of REV B amplified products using restriction enzyme digestion to semi-purify fragments

Since the specificity of REV B amplified products could not be determined through the DNA sequencing of the MGE-PCR amplified products, a different strategy for sequencing them was required. Since Rev B is a single primer that is designed against part of the mobile genetic element and appears at both ends, any PCR

amplified product would have that sequence at each end. This can potentially create problems with the sequencing reaction as forward and reverse sequences can overlap and this could explain the mixed peak sequences shown in Figure 4-5. Fragments can be generated by amplification between elements (see Figure 4-6A) or from a single element by means of mis-priming in the flanking region of the element (see Figure 4-6B). Differences in the position of the elements in different trypanosome stocks will generate different banding patterns. Thus we hypothesised that the generated bands from REV B amplification contained a RIME element sequence, and by cutting (digesting) the bands into smaller and specific sizes this will lead to successful DNA sequencing by single primer (REV B), whereby REV B will only have to sequence the digested DNA rather than the whole region of the amplified DNA.

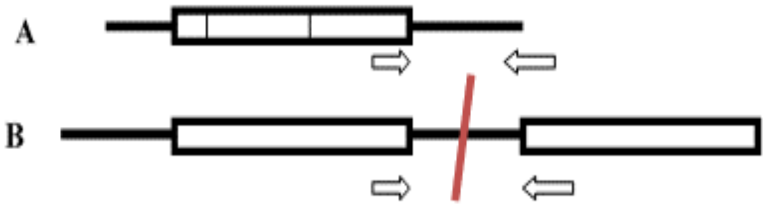


Figure 4-6. Schematic representation of MGE-PCR using the REV B primer to amplify the mobile element RIME. The red line in B indicates the restriction enzyme digestion site.

The sequence of the RIME element was put into a computer program (NEB Cutter http://nc2.neb.com/NEBcutter2/) to predict suitable restriction enzymes to cut the amplified fragment into two. Initially the digestion of the post REV B amplification products of sample 31 was carried out using MboI restriction enzyme. However, MboI failed to digest the REV B amplified products even with an extended incubation period overnight.

MboI failed to digest the MGE-PCR fragments that were produced by REV B amplification at two different incubation periods. The REV B amplified products failed to show up on the gel image, which could be due to there not being enough purified DNA, or could be due to the long period of incubation during which the DNA might have been degraded. Thus, at this point we cannot be sure if there were issues related to the enzyme and its components such as the buffer.

However, a different restriction enzyme was then used (AlwNI) to cut the post REV B amplification products. After overnight digestion in the water bath at 37ºC., AlwNI produced 1 cut (i.e. 2 bands) (see Figure 4-7) for all the selected bands except band 4 from sample 31.

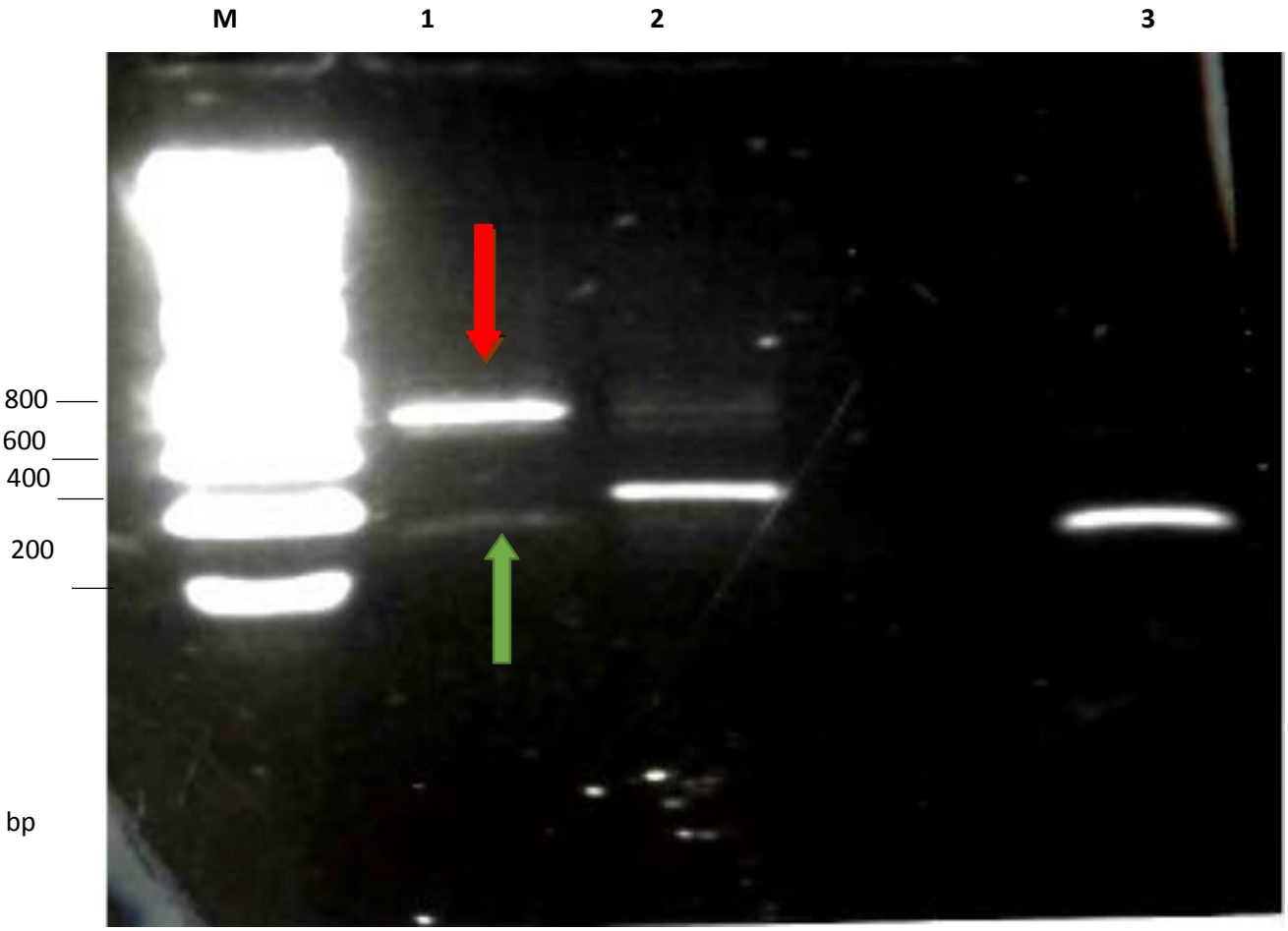


Figure 4-7. Restriction enzyme (AlwNI) digestion of MGE-PCR products. Overnight digestion by AlwNI for bands 1, 2, and 4 from Angolan sample 31 (highlighted in white arrows Figure 4-4). Lanes 1, 2 and 3 represent bands 1, 2 and 4 respectively. Bands 1 and 2 were successfully digested whereas band 4 failed to be digested. (M) represent the 1kb Bioline DNA marker. Red and green arrows indicate the selected bands for subsequent DNA sequencing

Subsequent to the successful digestion of the MGE PCR products by AlwNI, the digested bands from lane 1 (highlighted in red and green arrows respectively) were cut from the agarose gel and purified to be sequenced. The purified DNA was sent to Source Bioscience in Nottingham, UK for sequencing.

The sequencing data was analysed using the program FINCH TV. However, the DNA sequencing results of the digested bands produced still yet unreadable results (see Figures 4-8 and 4-9)

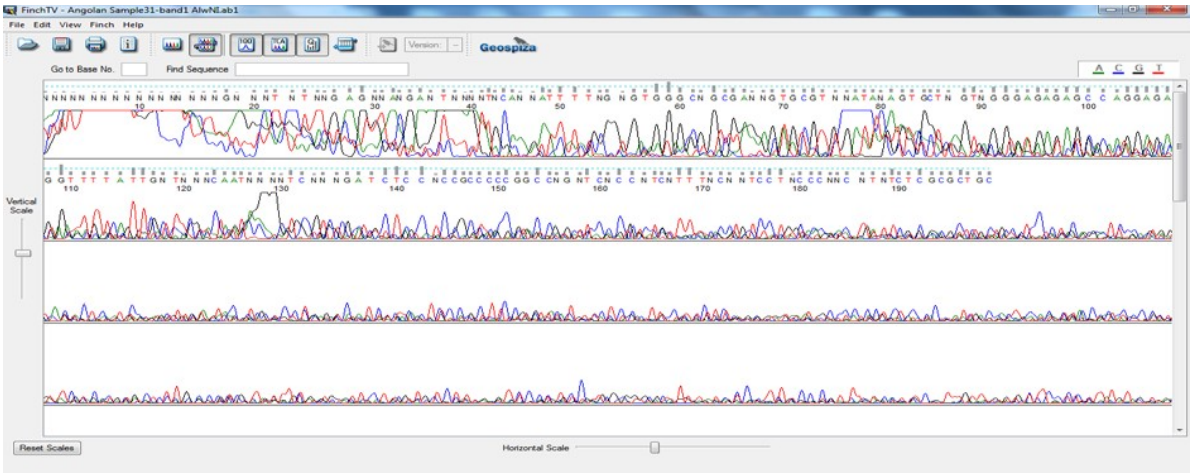


Figure 4-8. The DNA sequence of the AlwNI digested band (highlighted by the red arrow in Figure 4-7) by. The sequence data is unreadable.

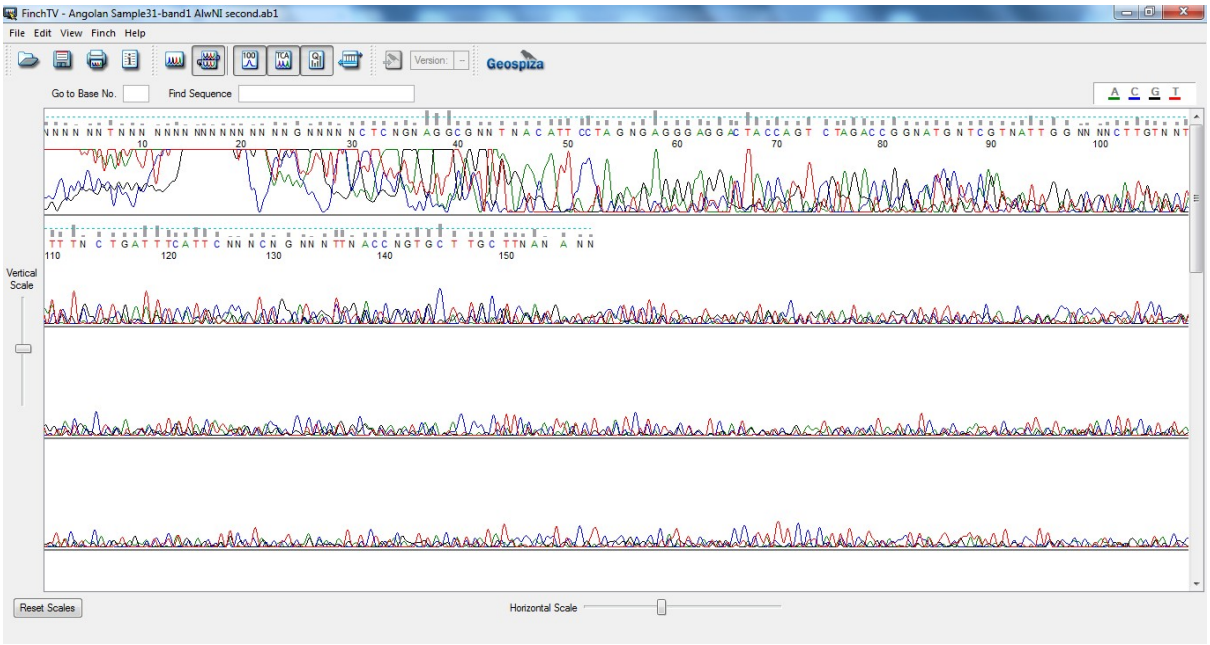


Figure 4-9. The DNA sequence of the AlwNI digested band (highlighted by the green arrow in Figure 4-7) . The sequence data is unreadable.

As can be seen from Figures 4-8 and 4-9 the results of the post sequencing of the digested REV B amplified products did not record any sign of improvement from the initial REV B sequencing results (as in Figure 4-5). The nucleotides of both bands were blasted in NCBI database to determine whether they were related to *T.brucei* or not, and the outcome was that the nucleotides were not significantly similar to any of the organisms present in NCBI.

4.4.5 Negative control amplification by REV B

Due to the failure of the DNA sequencing of the amplified products by MGE-PCR using the REV B primer, REV B was then tested on negative blood in order to determine whether REV B is only specific to trypanosomes. Control, uninfected, blood (fingertip) was spotted on a FTA card; following the treatment of the FTA card, DNA Chelex extraction was carried out on the punches taken from the FTA card. Then the eluted DNA was amplified with REV B. However, after the amplification REV B

appeared to not be a specific primer to trypanosome, as it managed to amplify the negative blood sample (see Figure 4-10). Thus a further investigation into mobile elements was required, in the interests of designing a primer that was specific to trypanosomes.

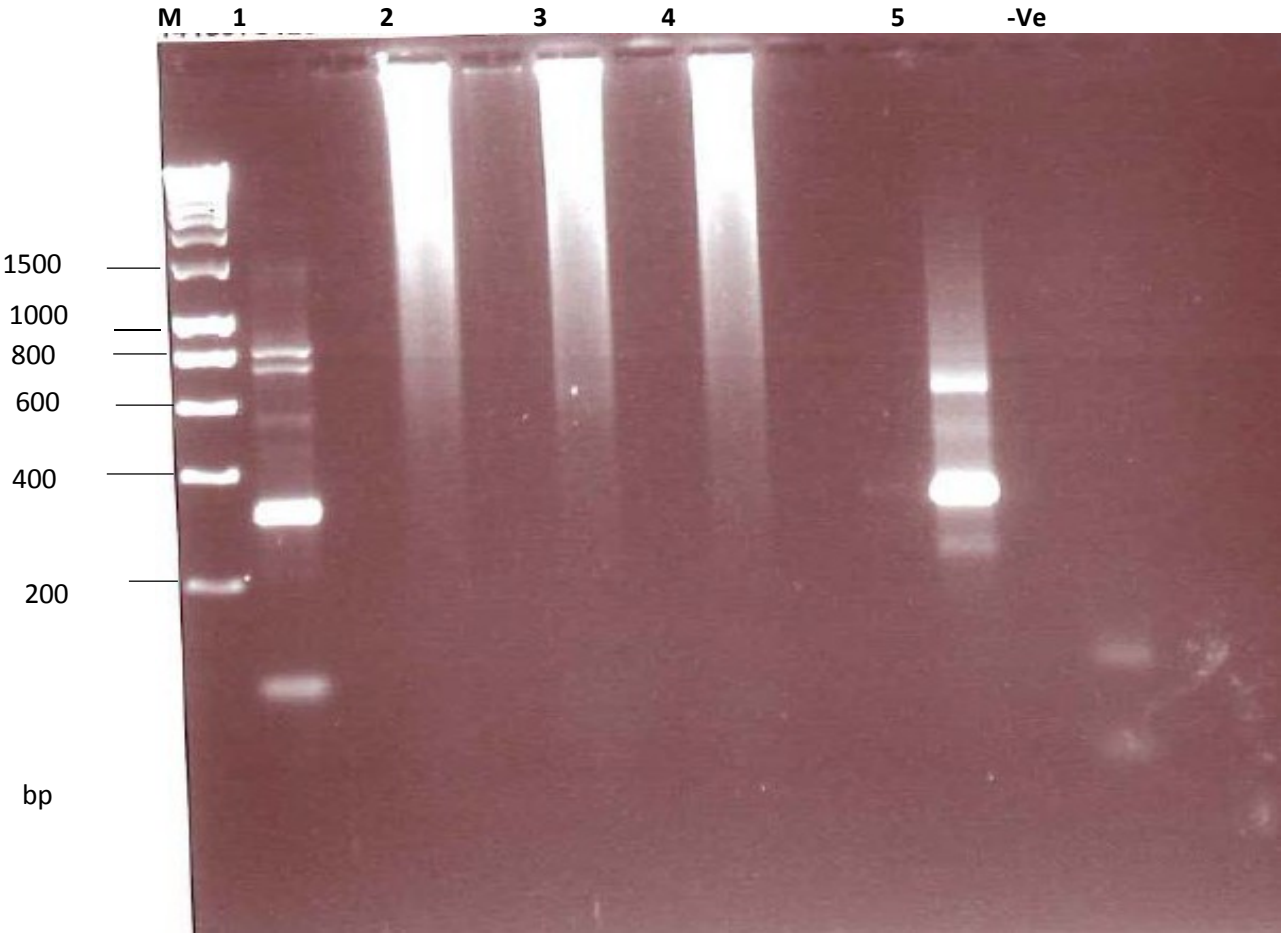


Figure 4-10. REV B amplification of a negative blood control. Lane 1 represents trypanosome DNA (positive control), lane 2 represents he FTA card punch spotted with trypanosome DNA, lanes 3 and 4 are Angolan samples (8 and 9 respectively), lane 5 is trypanosome-free blood and -Ve is the negative control (water).

4.4.6 Alternative approach to improve DNA sequencing data

Since there had been a failure of the sequencing of the REV B amplified products, including after the digestion of REV B amplified products by AlwNI, a different approach was carried out. The strategy here was to amplify the length variation of RIME element from the Angolan patient FTA cards using both by RIME A and RIME B primers, in order to obtain improved DNA sequencing data. This amplification

product will provide DNA sequences for a larger region of the gene, providing more precise sequences. It is more specific as the amplification products can be sequenced with different primers from each end.

The results show that out of the 36 Angolan samples only 7 samples were amplified which were 6, 7, 8, 9, 10, 12 and 31 (see Figures 4-11 and 4-12).

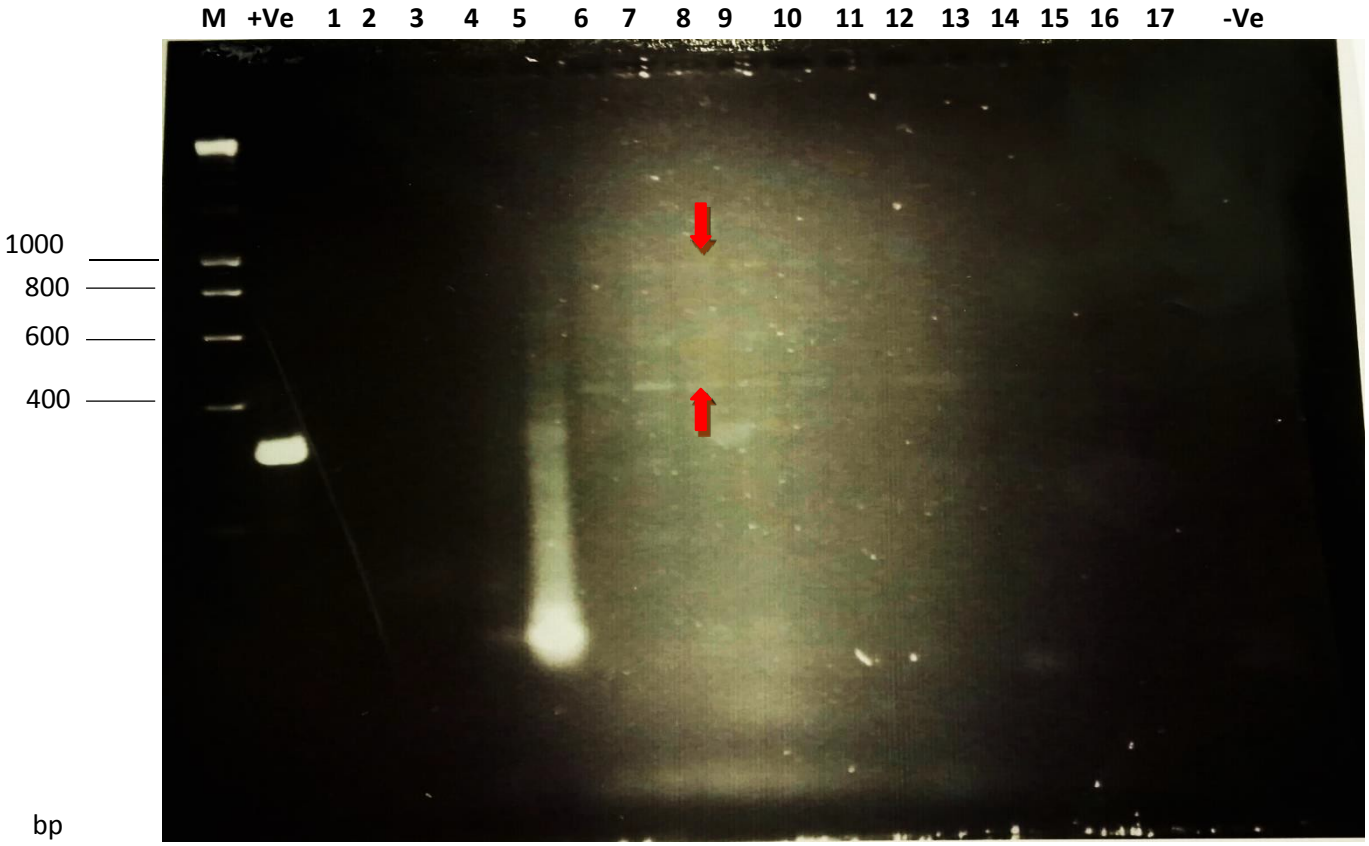


Figure 4-11. RIME A and B amplification of Angolan samples1-17. All the amplified samples share the same banding pattern, and this pattern is different from the positive control banding pattern (+ve). (M) represents the1kb Bioline DNA marker, +ve represents thepositive control (*T.b.rhodiense* DNA), lanes 1 to 17 represent the number of the Angolan samples and –ve represents the negative control (water). Note, the low concentration of the DNA has an impact of the banding densities. Red arrows highlight the selected bands for DNA sequencing.

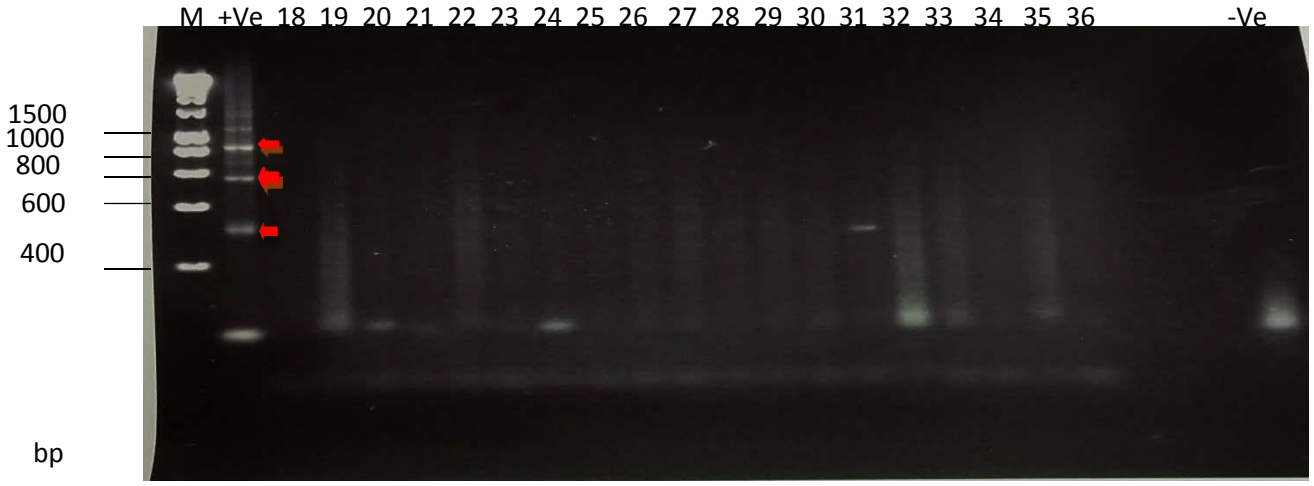


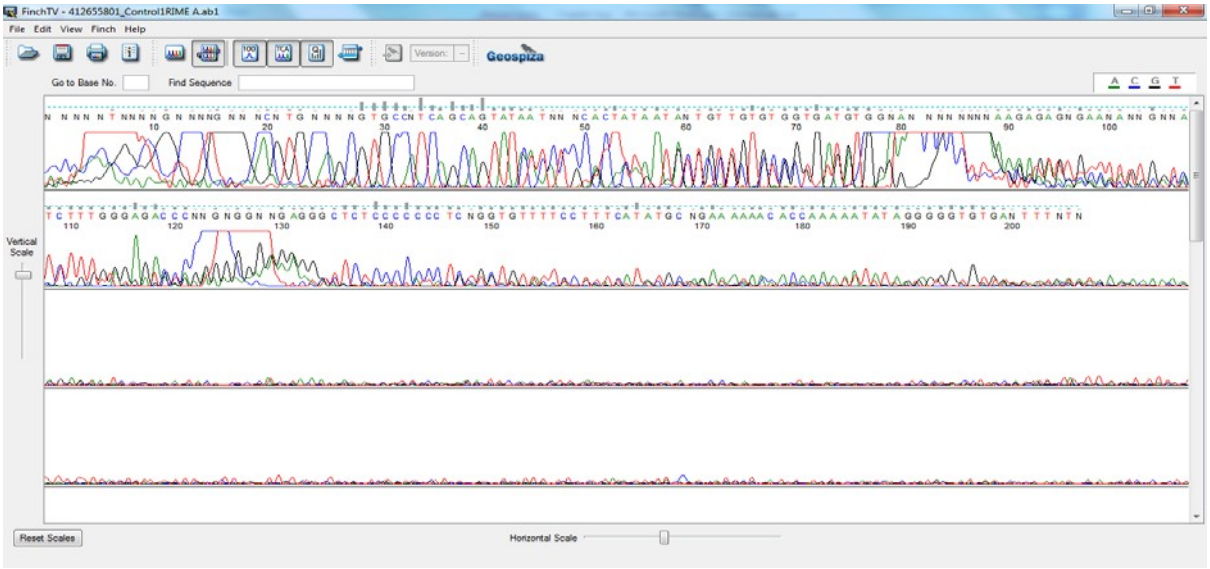
Figure 4-12. RIME A and B amplification of Angolan samples 18-36. (M) represents the 1 Kb Bioline DNA marker, lanes 18 to 36 represent the Angolan samples, +ve represent the positive control (*T.b.rhodiense* DNA) and –ve represent the negative control (water). Red arrows highlightthe selected bands for DNA sequencing.

4.4.7 Post RIME A and RIME B amplification analysis (DNA sequencing)

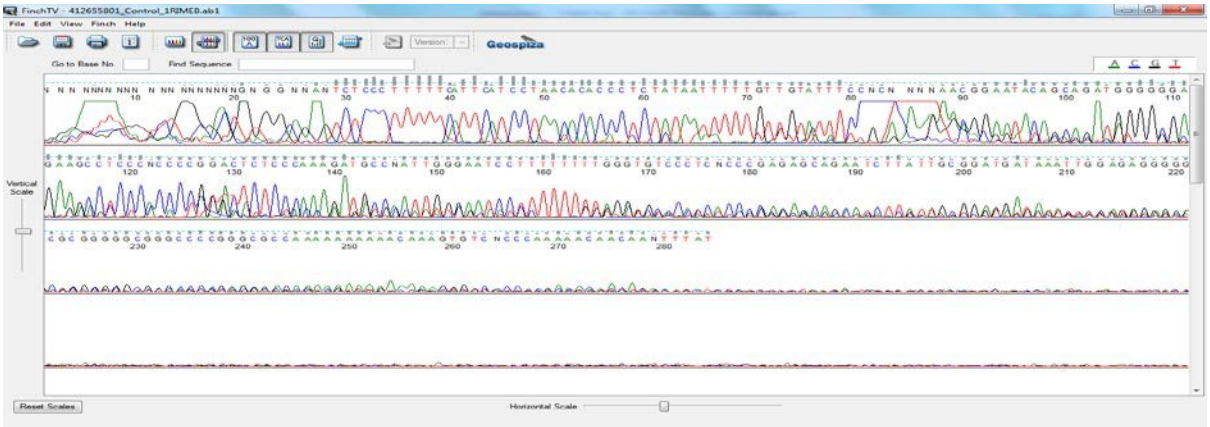
To study the amplicons generated by RIME A and B amplification, the amplified bands were cut and purified and sent for sequencing. Bands that were marked with red arrows including the positive control from Figure 4-12 and the bands (marked with red arrows) generated from the amplification of the Angolan samples (sample 8 Figure 4-11) were selected for DNA sequencing. The DNA sequencing was carried out by Source Bioscience in Nottingham (UK).

After obtaining the DNA sequence data, the data was analysed using Finch TV software. However, unfortunately the results yet again failed to be readable. (Figures 4-13 and 4-14) show the screen shots of the DNA sequences of the positive control (*T.b.rhodesiense*DNA) and the amplified Angolan sample 8 respectively using Finch TV software; the screen shots are arranged in accordance with the molecular weight of each band (from the highest molecular weight to the lowest).

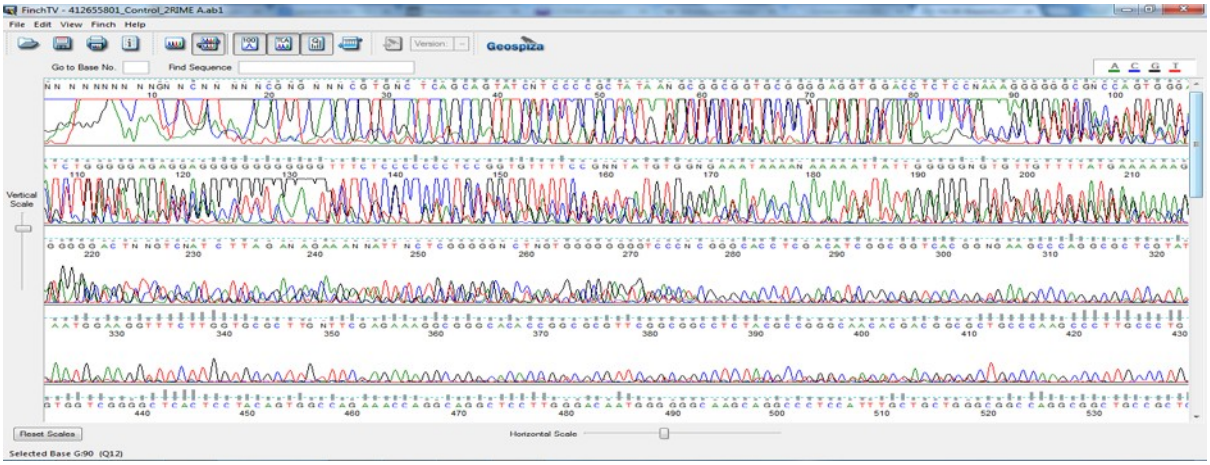
Figure 4-13. Six screen shots of DNA sequences of RIME A and RIME B amplification products (see Figure 4-12).



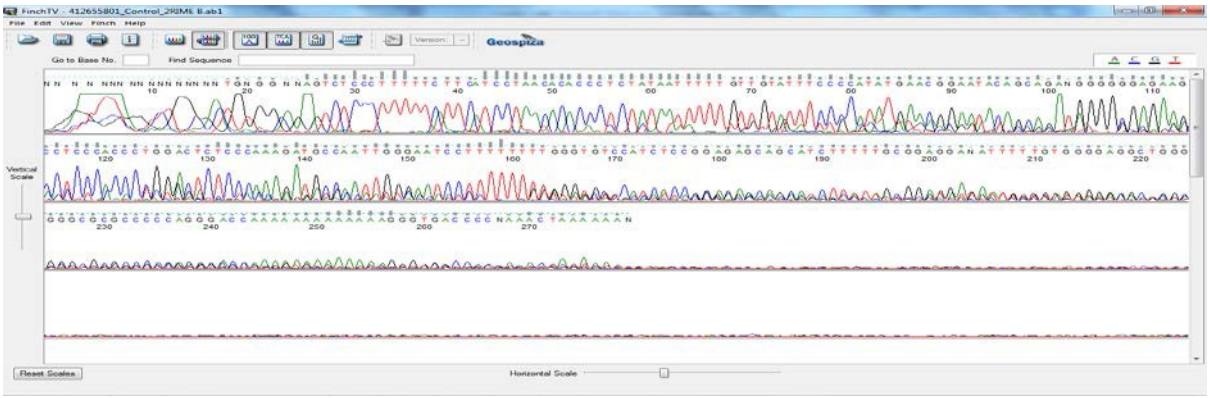
The above diagram illustrates the sequencing of positive control sample DNA by RIME A. According to the molecular weight of this band which is 1000bp only 200bp have been sequenced, and the peaks of the sequencing are mixed which has led to unreadable results.



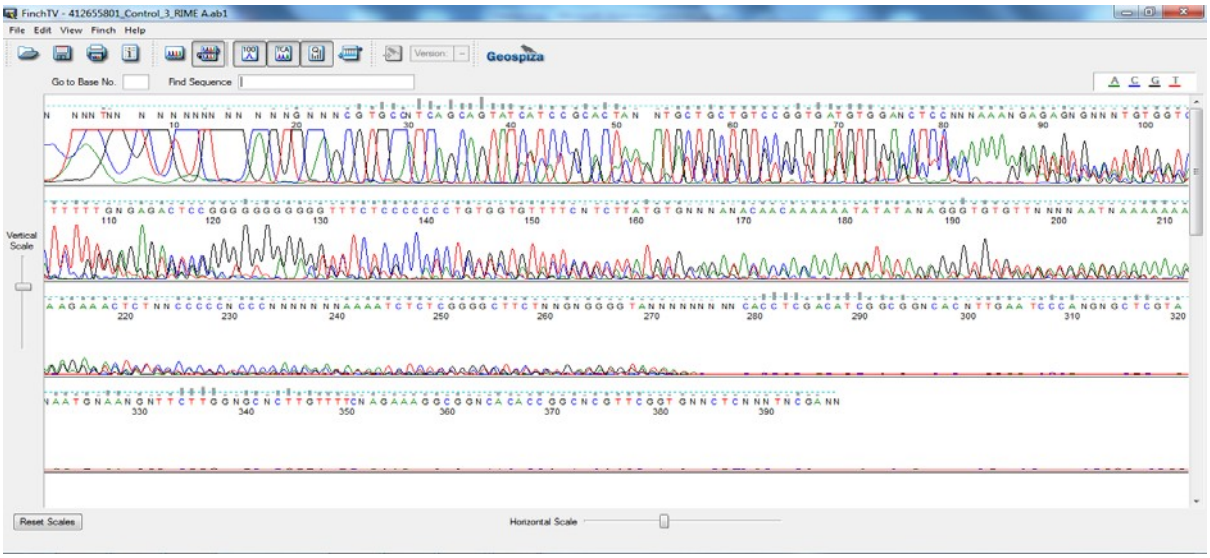
The above diagram shows the sequencing of positive control sample (trypanosome DNA) by RIME B. The second band molecular weight is estimated to be 1000bp, and again only 200bp have been sequenced. Due to this incomplete and bad sequence alignment between RIME A and Reverse RIME B was not possible.



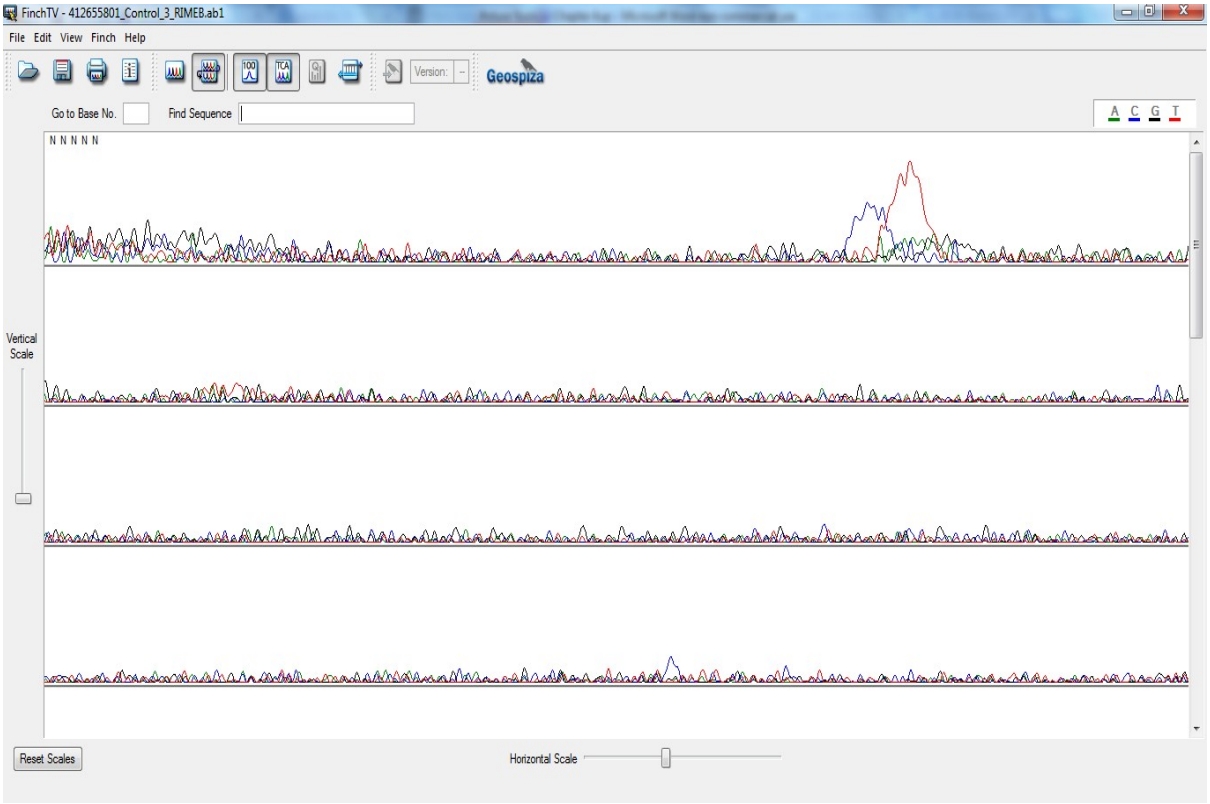
The above diagram shows the DNA sequencing of the second highest molecular weight band of positive control sample; the sequencing was carried with RIME A primer. The estimated molecular weight of this band is 800bp.



The above diagram shows the DNA sequencing of the second highest molecular weight band of positive control sample; the sequencing was carried with RIME B primer. The estimated molecular weight of this band is 800bp.

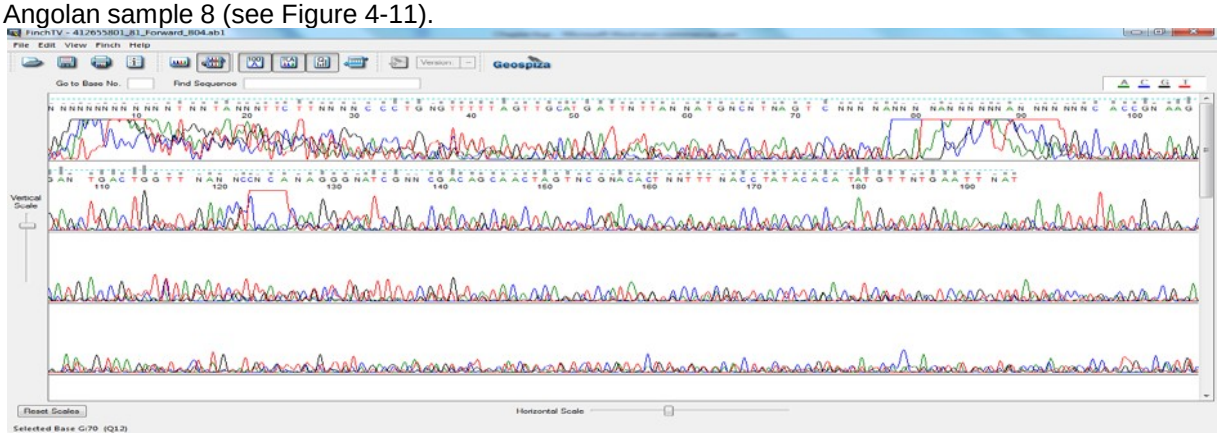


The above diagram illustrates the DNA sequencing of the third largest (molecular weight) band of the positive control sample. RIME A was used in this sequencing, the estimated band size is 600bp. Only around 200bp that have been "sequenced".

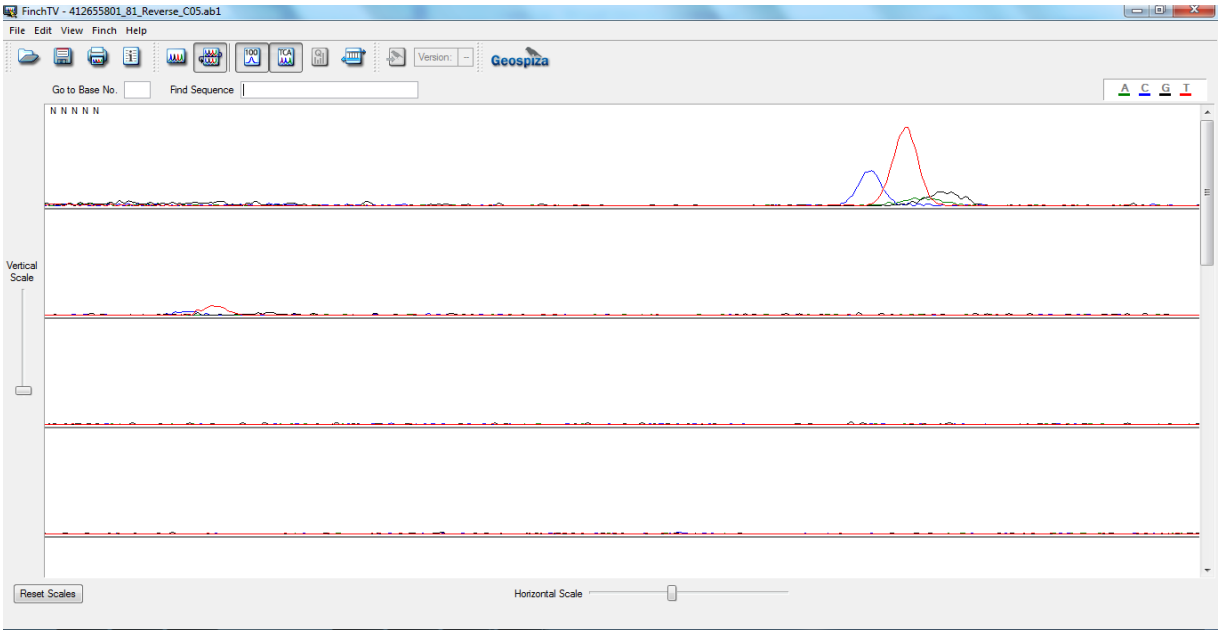


The above diagram shows the DNA sequencing of the RIME B amplification of the third band of the positive control sample from (see Figure 4-12). However, it failed to result in good sequencing.

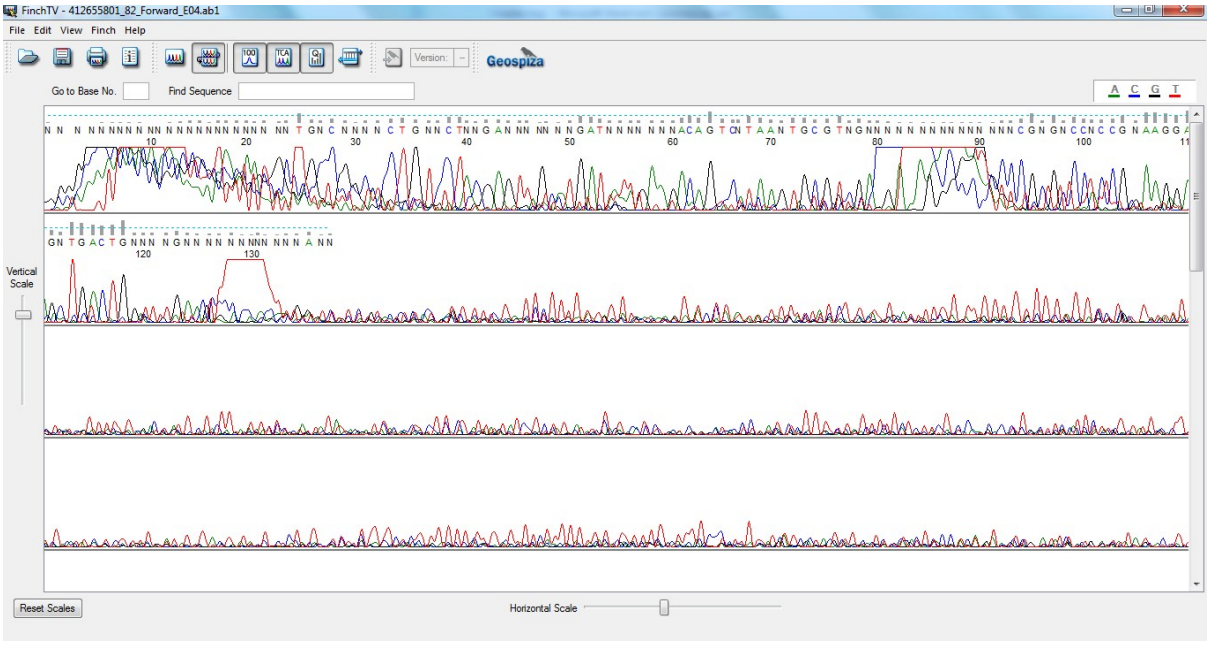
Figure 4-14. Four screen shots of the DNA sequencing of RIME A and RIME B PCR products from Angolan sample 8 (see Figure 4-11).



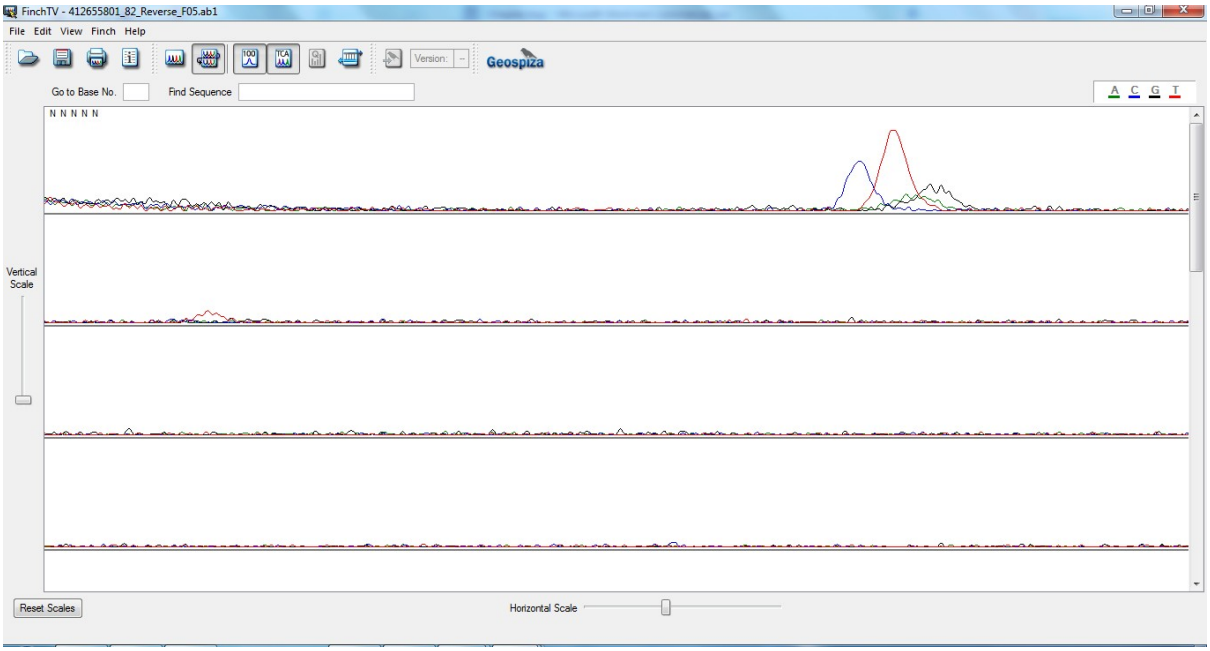
The above diagram illustrates the DNA sequencing of the first band of the Angolan sample no. 8. This sequencing was carried using RIME A primer, and again the sequencing is relatively similar to the sequencing of the positive control in term of the low quality.



The above diagram illustrates the DNA sequencing of the first band of the Angolan sample 8. This sequencing was carried by using RIME B primer; again the sequencing is very badly sequenced with a very low quality.



This diagram (above) illustrates the DNA sequencing of the second band of the Angolan sample 8. This sequencing was carried using RIME A primer, again the sequencing is relatively similar to the sequencing of the positive control in terms of the low quality.



The above diagram illustrates the DNA sequencing of the second band of the Angolan sample no. 8. This sequencing was carried using RIME B primer, and the quality of the sequencing is poor.

According to the above screen shots of the DNA sequencing by RIME A and RIME B, the quality DNA sequences of either the positive control or the Angolan sample is very poor. The extraction protocol was nonetheless carried out carefully in accordance with the booklet for the extraction kit. Also, before sending the purified DNA to the sequencing company, the concentration of the purified DNA was measured by a nanodrop spectrophotometer at 280nm, to comply with the requirements that are set by Source Bioscience. Overall the sequences have recorded a sign of small improvement particularly in the control sample amplified DNA. The sequences of readable sections (i.e. good quality of nucleotides run 10-70bp) were subject to a BLAST search to attempt to identify them. The resultant organisms were completely irrelevant to trypanosome. The conclusion is that it was not trypanosome DNA that was amplified. These poor results could be due to several reasons including the following: low template concentration - particularly in the case of the Angolan samples, the template may not be completely purified so it still carries unincorporated nucleotide such as dNTP or excessive agarose gel; or nonspecific primers can cause the primers to bind to more than one site and the resultant readings are overlapped. To rule out the possibility of the low template concentration, the Angolan sample no. 8 was re-amplified with RIME A and RIME B
in order to increase the concentration of the amplified DNA. Yet the DNA sequencing resultant readings were still of poor quality (data not shown).

4.5 Discussion

Recently, the World Health Organisation (2012) has established a road map on neglected tropical diseases, especially focused on the elimination of human African trypanosomiasis as a public-health problem by 2020. Thus, a better understanding of epidemiological principles and the development of new control and surveillance tools will certainly help in achieving the goal of elimination.

In this present study one of the molecular tools (MGE-PCR) used to study and detect the trypanosome has been investigated in great detail, in order to compare the two detection methods - ITS-PCR (Chapter 3) and MGE PCR - in terms of specificity and sensitivity. Previously, MGE-PCR has been proven to be specific in differentiating between trypanosome strains when using purified DNA. However, to our knowledge MGE-PCR has not been applied on field samples that contain host DNA applied to FTA cards. Thus, MGE-PCR was carried out on DNA extracted using a Chelex extraction protocol from 36 blood samples spotted onto FTA cards which were collected from Angola that were infected with *T.brucei* as tested by CATT.

Initially a single primer REV B was used to amplify the samples. However, the initial concentration of the primer (10pmol) had a negative impact on the amplification, since the samples including the positive control failed to amplify. Czerny (1995) has suggested that the primers have to compete with the accumulating product in finding their target sequence, which could lead to limiting the reaction at late cycles. He tested whether the ratio between primer and product concentration influences the product yield as he performed reactions with increasing concentrations of primers but in otherwise identical conditions. He observed an increase from between 2- and

almost 20-fold depending on variation in primer concentration from 1 to 20 pmol. Therefore, the concentration of the primer (REV B) was increased to 20pmol, which led to a successful amplification of the samples.

In term of sensitivity MGE-PCR appeared to record a higher degree of sensitivity than the nested ITS-PCR (Figure 4-3 and 4-4), with 58.3% (21 samples) have been amplified. Single primer PCR allows amplification from both known and unknown regions of the DNA and this means that there is a considerably greater chance of non-specific amplification. An initial indicator that the MGE-PCR was behaving non-specifically came from the fact that the samples were collected from different parts of Angola and yet they shared the same banding patterns (although they did differ from the positive control).

It was clear from the DNA sequencing that specific amplification products were not being obtained. Thus the REV B primer was behaving non-specifically within the MGE-PCR when in the presence of host DNA on the FTA cards. The quality of the sequencing of the amplified products was poor. This could be due to different reasons including: (i) non-specificity in the primer, or (ii) the low concentration of DNA yielded after the amplification, as can be noticed from Figures 3 and 4 where the amplification has resulted in smears. Also the number of PCR cycles plays an important role in generating the number of copies produced after the amplification, so another factor is (iii) the purity of the DNA before it is sent for sequencing, as well as (iv) the age of the sample as the samples are believed to have been collected in the 2001. One common feature of ancient DNA samples is the presence of miscoding lesions which cause the incorporation of incorrect nucleotides during DNA

amplification. In 2001, Hofreiter *et al.* showed for the first time that most of these damage-derived errors are caused by hydrolytic deamination of cytosine into uracil leading to apparent C to T or G to A substitutions in DNA templates sequenced after PCR amplification. A further factor could be (v) the extraction protocol that was used to extract the template DNA prior MGE-PCR. Mseddi *et al.* (2001) have reported a comparative study between Chelex extraction and urea Chelex extraction on *Cryptococcus* samples that were collected between 2005 and 2009. When they had incubated the samples in 2ml of urea buffer for 3h the latter protocol of extraction gave a higher DNA concentration (16,03 µg/ µl) than the Chelex extraction (7.54 µg/ µl).

A different approach was thus followed, which involved digesting a selected band from the agarose gel using AlwNI, which cut the amplified DNA into specific sized pieces that would result in the primer sequence only being on one end of the amplicon thus reducing the chances of mis-priming and the generation of a false sequence. However, yet again the quality of the DNA sequencing was poor. A final attempt was made to study the specificity of REV B since the DNA sequencing failed to produce any acceptable results; the REV B was used to amplify a negative blood (uninfected blood) sample to determine whether it would amplify it or not. Unfortunately REV B appears to be nonspecific to trypanosomes when used on FTA cards as the negative control blood was clearly amplified and produced relatively similar banding pattern that was produced by trypanosome DNA.

Finally, due to the obstacles relating to single primer PCR (REV B) mentioned above, another protocol was carried out on the Angolan samples, involving amplifying them with forward and reverse primers (RIME A and RIME B). This could increase the number of copies produced post amplification, and determine whether

the age of the samples have had an impact on the poor quality sequencing results generated from REV B. Firstly the number of the amplified samples was less (8 samples) than with the REV B amplification. After the amplification the specificity of RIME A and RIME B primers have been studied by sequencing the amplified samples including the positive control. Yet the quality of the produced sequencing results still poor; this could be due to two or more templates being present in the reaction, the PCR fragment not being purified of leftover primers before sequencing or poor quality PCR templates containing multiple DNA fragments being used. On the other hand, in the positive control sample where short DNA sequencing read lengths are observed this could be due to too much template DNA, or too much primer.

In summary, the development of control and genotyping tools for HAT plays an important role in efforts aimed at eliminating the disease. MGE-PCR has in the past proved to be one of the most powerful tools that can be used in studying the epidemiological aspects of HAT. For the first time MGE-PCR has been carried out directly on a field sample and it shows potential although currently the technique appears to be behaving non-specifically. Further study is required into the behaviour of mobile elements and their properties for the purpose of strengthening the use of MGE-PCR in terms of specificity as a molecular epidemiological tool.

Chapter 5 Investigation of Mobile Genetic Elements Found Inserted in Retrotransposon Hotspot Genes (*RHS*) in the *T. brucei* TREU927 Genome

5.1 Introduction

As discussed in Chapter 4 mobile genetic elements can be used as markers to study the epidemiology of sleeping sickness. However, MGE-PCR cannot be used to provide data on samples that contain mixed genotypes, because traditional DNA sequencing using a single primer that recognises mixed sequences will return unreadable sequences. The identification of these sequences is also not possible because the Rev B single primer will be found at both ends of the PCR product. Thus, a further and more detailed investigation into the mobile gene elements is required to better understand their role and behaviour and in order to overcome the sequencing challenges by developing a marker based on the MGEs and a single PCR primer.

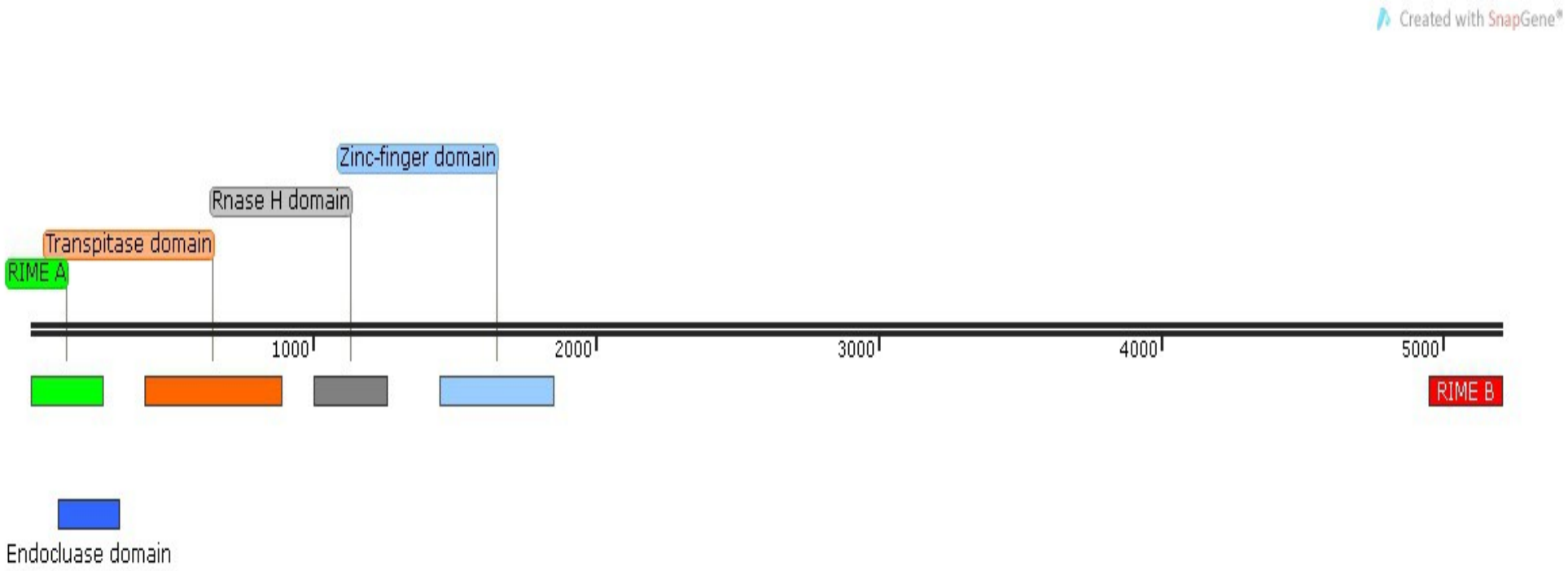
Repetitive sequences can be artificially divided into two groups: interspersed repeats and tandemly repeated DNA. Interspersed repeats mainly represent inactive copies of presently or historically active transposable elements (Wickstead *et al.,* 2003). MGEs are also known as transposable or retrotransposons elements (TEs) due to their mechanism of mobilisation (Bringaud *et al.,* 2008). TEs are defined as 'parasitic' genetic elements that are spread by replicating themselves within the host genome (Kofler *et al.,* 2012). TEs were first described in 1950 by Barbara McClintock, and she suggested that although TEs are considered to be junk DNA they play a significant role in genome evolution (Piskurek and Jackson, 2012). However, there is now a growing body of evidence which suggests that TEs may be functionally important and not just ‘‘junk,’’ ‘‘selfish,’’ or ‘‘parasitic’’ DNA sequences that make as many copies of themselves as possible (Biemont and Vieira, 2006). For example,

there is a considerable number of domesticated TE copies which act as transcriptional regulatory elements or contribute to the protein coding of cellular genes (Shapiro, 2005). They can contribute to genome plasticity and to genome architecture. For example, the mosquitos *Aedes aegypti* and *Anopheles gambiae* have genomes significantly larger than the model organism *Drosophila*. Most of this genome increase is due to the high proportional content of repetitive DNA and transposable DNA. This feature can be used to study genome evolution, shaping both genes and genomes in terms of function and structure (Blaxter *et al.*, 2003). Also, since most MGEs are non-autonomous, since they exist in large number, and since their transfer usually occurs vertically, they can shape the evolution of distant parts of the genome. Mobile elements are involved in creating mutations and genomic rearrangements and, in many eukaryotes, these effects can be regulated through a RNA silencing mechanism such as RNA interference (RNAi) (Girard and Hannon, 2008). In addition, based on the mobilisation mechanisms, TEs can be classified into two major groups: firstly, those that move via reverse transcription using an RNA intermediate and are known as retrotransposons, or class I. Two to five percent of the trypanosomatid genome contains class I elements. The second class is known as DNA transposons or class II, and they move using a DNA intermediary. Approximately 65% of the human genome contains class II TEs (Thomas *et al*., 2010). In addition, class I is further subdivided into long terminal repeat retrotransposons (LTR) and non LTR retrotransposons (Bringaud *et al*., 2008)

Since the genome of *T.brucei* was sequenced in 2005 (Berriman *et al*., 2005) much more information is known about these. It is reported that the trypanosome genome contains hundreds of copies of apparently dead retrotransposable elements and probable reverse transcriptase pseudogenes. Many copies of these elements

appear to be interspersed with VSG genes (Cross, 2003). The Non-LTR retrotransposons are found to be the most abundant mobile elements described in the genome of *T.brucei* (RIME, *ingi* and SLACS) (Hasan *et al.*, 1984; Kimmel *et al*., 1987; Aksoy, 1991). Kimmel *et al.* (1987) reported *ingi* to be randomly distributed in the *T.brucei* genome. The *ingi* is 5.25 kb and is characterized by its autonomous features. RIME is 0.5 kb and appears to be truncated version of *ingi*. The ends of the *ingi* are bordered by two separate fragments of RIME (RIME A 0.25kb and RIME B 0.25kb), at the 5' and 3' ends respectively (see Figure 5-1). RIME elements are short non-autonomous reteroelements (Bringaud *et al*., 2004). Moreover, it is estimated that the *T*.*brucei* genome contains ~400 copies of *ingi* that is approximately equal to 5% of the genome, and the RIME is estimated as being at 86 copies per haploid genome of *T.brucei* (Donelson *et al.*, 1996; Bringaud *et al.,* 2002). Only 2.5% are considered as potentially active (Bringaud *et al.,* 2008).Furthermore, *Ingi*/RIME are only located in the megabase chromosomes. In addition, *ingi*/RIME are found at the end of telomeres as they proceed or are within most of the expression sites for the VSG genes. In addition, it is still unclear whether either *ingi* or RIME gave rise to one or the other and we do not fully understand the mechanism by which this may have occurred (Bhattacharya *et al*., 2002).

Figure 5-1. A schematic illustration of ingi (Accession M16068.1) and RIME elements. . Functional domains of the ingi are colour coded as following: endonuclease domain transcriptase domain (orange), Rnase H domain (grey) and Zinc finger domain (light blue). The 0.5kb RIME flanked on the 5' and 3' ends of the 5.25kb ingi are colour coded as red and green for RIME B and A respectively. Diagram was constructed by Snap Gene 3.43



RIME has proven to be a marker for differentiating between individual strains from species of trypanosomes collected from different geographical locations (Tilley *et al.,* 2003). However, in this present study the use of RIME in differentiating between two *sub*-species will be investigated. Wickstead *et al.* (2003) have reported that no two organisms can share a common mobile element with the exception of *T.brucei* and *T.cruzi*. However, the trypanosomal mobile elements in the two species vary significantly in genomic location. Furthermore, the diversification characteristics of the mobile element can be entirely restricted to a species such as *TARE-2* and *TARE-3* elements in *P. falciparum* (Figueiredo *et al.,* 2000). Thus, it can be concluded that MGEs possess evolutionary features that could enable them to be used as tools to track the spread of infections by following particular evolutionary variants through populations of parasites. Also, the relatively small size of RIME (0.5kb) can make RIME a useful marker in differentiating between *T.brucei* sub-species.

Although, *ingi*/RIME are previously thought to have been distributed randomly around the genome, Bringaud *et al.* (2002) reported that the distribution of ingi/RIME are associated with a number of highly repeated gene families including rRNA genes, tandemly repeated tubulin gene arrays and retrotransposon hot spot (*RHS*) genes.

RHS genes are divided into six sub-families (*RHS*1-6) on the basis of the divergent C-terminal domain of their gene product. *RHS* genes are highly repetitive multigene families and are composed of about 280 copies per diploid genome of the *T.brucei*. About 60% of them are non-functional pseudogenes. The *RHS* (pseudo) genes appear to be frequently located in the subtelomeres adjacent to VSG expression sites (Bringaud *et al.,* 2002). Moreover, Bringaud *et al.* (2002) have reported that about one-third of *RHS* pseudo genes contains ingi/RIME inserted mobile elements inserted in the relatively same position in the *RHS* genes usually at a 12-mer sequence

(TACTGTTATACA).

The relationships between taxa in the *T. brucei* complex are not fully understood (Koffi *et al.,* 2009; Balmer *et al*., 2011). The current subdivision of *T. brucei* is based on host range, pathogenicity and geographic origin, but it does not reflect genetic or evolutionary distinctions (Gibson, 1986; Balmer *et al.,* 2011). However, the genomic variation between *T.b.brucei* and *T.b.gambiense* is only limited to less than 1%. This limited genomic variation suggests that the observed functional differences could be due to features shared between subspecies, which vary in either structure or expression (Jackson et al., 2010). *T. b. gambiense* strains have the smallest genomes in the *T. brucei* species complex, having 71-82% of the DNA content measured for *T. b. brucei* (Kanmogne *et al.,* 1997).

In this study a bioinformatics investigation will be carried out on the mobile elements that are found inserted in the *RHS* pseudogenes of the TREU 927 genome. This is in order to understand their behaviour and evolution with a view to discerning how they can be used as differentiating markers between different *T.brucei* species and strains.

5.2 Objectives

- To comparatively investigate the genomes of *T.b.brucei* and *T.b.gambiense* to determine the presence and absence of MGEs inserted within the *RHS* genes.
- To identify *RHS* genes and their families present in the TREU 927 genome.
- To classify *RHS* genes via a phylogenetic analysis of *RHS* genes and an investigation into possible evolutionary time progression events in the *RHS* genes.

5.3 Methods

This study applied a series of bioinformatical protocols including: multiple sequence alignments using CLUSTAL W (EBI), and MEGA 5, designed to classify the *RHS* genes/pseudogenes according to their subfamilies. In addition, MEGA5 was used to build the phylogenetic tree of the *RHS* genes. Finally BLAST searches were used to identify the *RHS* genes' subfamilies, and to execute a retro-analysis comparing the *T.brucei* and *T.b.gambiense.*

Details of the bioinformatics methods are either included in Chapter 2, (program instructions) or, are applied, within the results section of this chapter.

5.4 Results

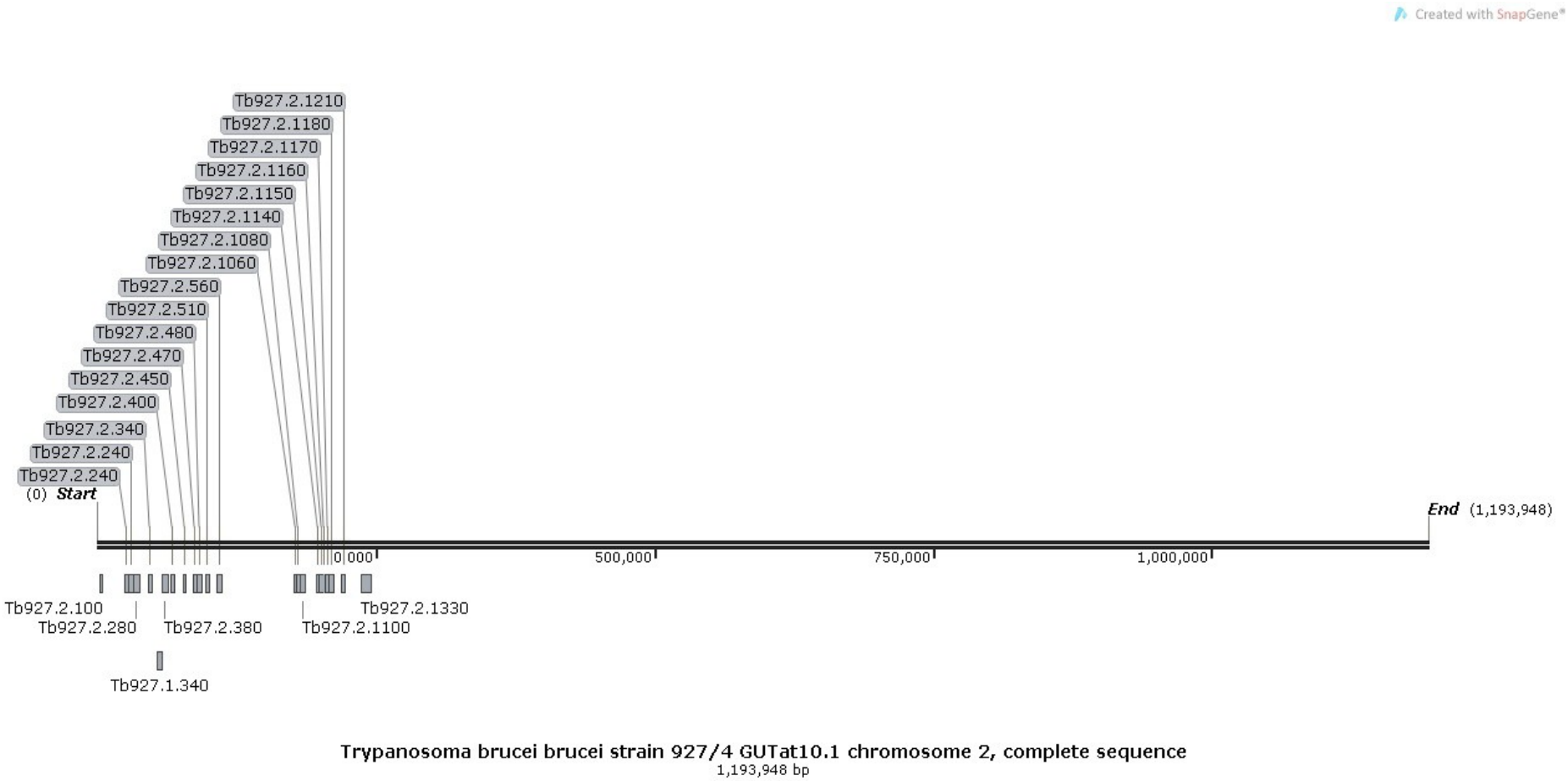
Since REV B MGE-PCR (see Chapter 4) failed to be specific enough as a primer/amplification system for the detection of trypanosomes when used directly on FTA cards. A bioinformatics approach was used to investigate more specific sequences (based on MGEs) that could be used for trypanosome molecular epidemiology. It is important to study the genetic background of mobile elements, in order to study their behaviour, their role as mobile elements and their potential for use as epidemiological markers. A sequence of bioinformatic “experiments” was undertaken to investigate the characteristics of mobile elements and their interaction with the *RHS* genes which function as hotspots for the insertion of mobile genetic elements.

5.4.1 Identification of retrotransposon hot spot (*RHS*) genes in the TREU927 genome

RHS genes are known as sub-telomeric genes, located at the telomere ends of the *T.brucei* chromosomes. Figure 5-2 shows an example of this for Chromosome 2. A search was conducted on NCBI map viewer using two terms - "*RHS*" and "retrotransposon hot spot" - and this resulted in 59 *RHS* genes being reported throughout the 11 chromosomes of the TREU927 strain. *RHS* genes were found in all chromosomes with the exception of chromosome 8. There are sub-families of each of the *RHS*/pseudogenes (*RHS* 1 – 6) mentioned in the database and an alignment was carried out on each *RHS* gene with a representative distribution of *RHS* gene families (1- 6) present in the database under the following accession numbers: AY046887 (*RHS*1), AY046888 (*RHS*2), AY046889 (*RHS*3), AY046890 (*RHS*4), AY046891 (*RHS*5), and AY046892 (*RHS*6) (Data not shown). Out of the 59 *RHS*

amino acid sequences there were 29 *RHS* genes and 24 pseudogenes. This contrast with the findings of Bringaud *et al.* (2002) who have reported the presence of *RHS* genes (21 copies) and pseudogenes (24 copies). The search resulted in 6 degenerate copies in which amino acid sequences did not belong to any of the *RHS* genes' sub- families. Degenerate copies were excluded from phylogenetic analysis (see below) due to their high divergence, with the exception of a *RHS* degenerate gene (Tb927.2.1150) which was included to see how it fitted into phylogenetic trees of the *RHS* gene family. Eleven of 59 *RHS* genes were interrupted by mobile element insertions (see Figure 5-3) and all were pseudogenes. Multiple alignments of the amino acid sequences for the retrieved *RHS*/pseudogenes were carried out using EBI clustalW2 and assembled for comparison (appendix 2)

Figure 5-2. RHS /pseudo genes found in chromosome 2 in the TREU927 genome using NCBI map viewer. *RHS* pseudo gene accession numbers are boxed in grey. The diagram was constructed with SnapGene 3.4.3.



In addition to the *RHS* 1-6 families, the analysis using NCBI map viewer showed that among the 59 *RHS* genes there were 8 *RHS* genes and 1 *RHS* pseudogene in chromosome 7 that do not belong to the *RHS* families as reported by Bringaud *et al.* (2006). These were defined as "unclassified" but all possessed the retroelement insertion target.

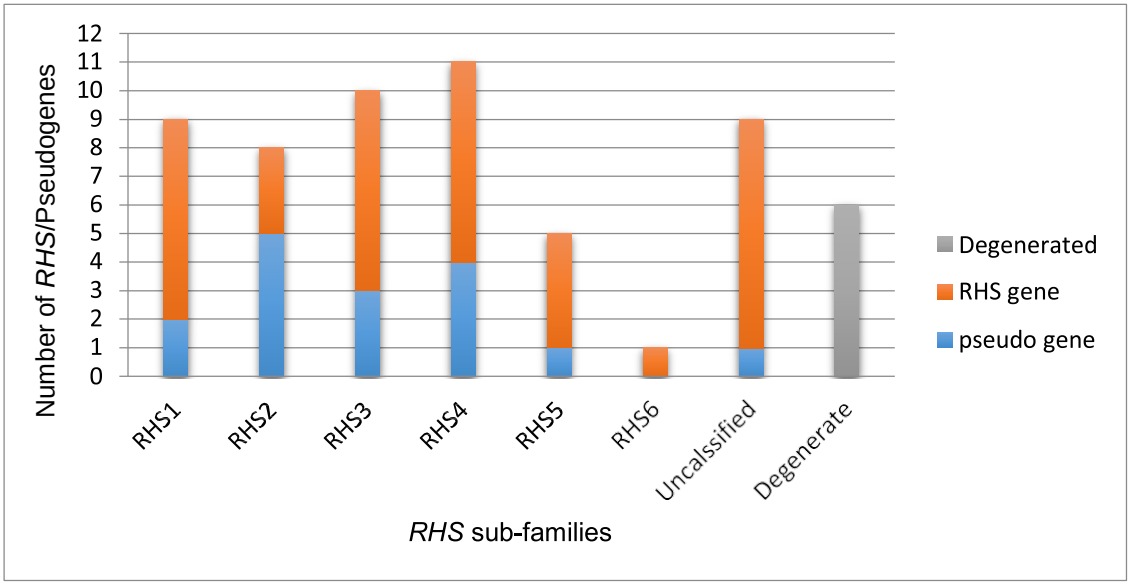


Figure 5-3. The number of the RHS genes according to their families in TREU927 genome. The most abundant *RHS*/pseudogene family is *RHS*4

5.4.2 Evolutionary relationship of *RHS* genes:

To study the relationship between the *RHS* genes present in the TREU927 genome, a phylogenetic tree (using a Maximum Likelihood algorithm) was constructed using the MEGA5 programme and based on the multiple alignment of amino acid sequences for each *RHS* gene which were retrieved from the GeneDB database and the reference *RHS* gene families (1-6) (see Figure 4). According to the phylogenetic tree there are two clades of *RHS* genes which include *RHS*3 / *RHS*4 and *RHS*5 / *RHS*6 respectively. These clades were formed based on the fact that an intergroup alignment of the *RHS*3/*RHS*4 is estimated to be 43%, whereas *RHS*5/6 is estimated to be at 50% (Bringaud *et al.,* 2002).

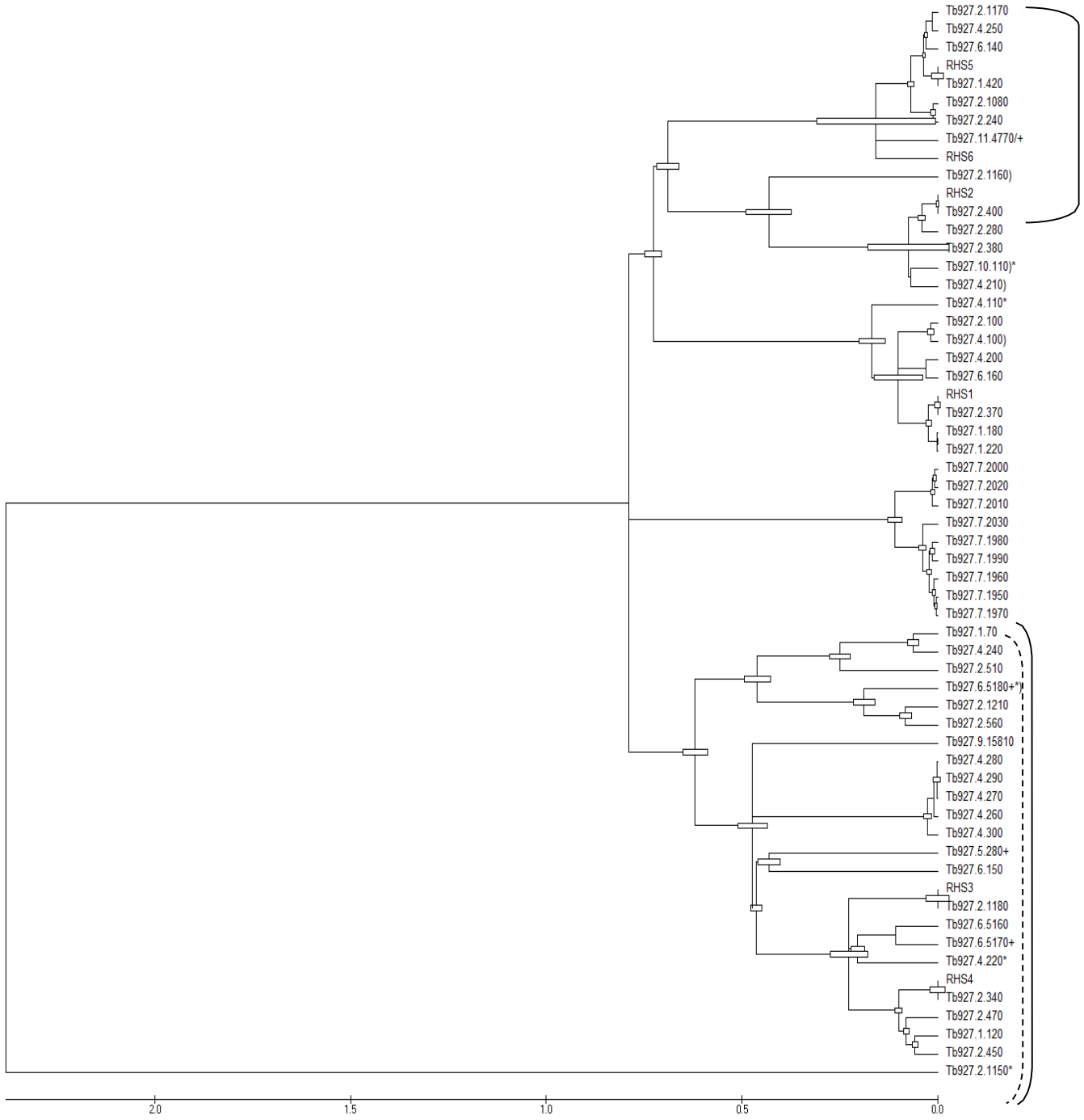


Figure 5-4. Molecular phylogenetic analysis of the TREU 927 RHS genes using the Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). *RHS* genes with mobile element insertions are marked with + for (truncated ingi), * (RIME A),) for (RIME B), and (/) for DIRE. The phylogenetic tree was obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood.

Due to the complexity involved in analysing all *RHS* genes and the relative abundance of subfamilies of *RHS*3 and *RHS*4 in TREU 927, only members of the *RHS*3 and 4 families were analysed in this here (highlighted by the dashed bracket in Figure 5-4). Before the investigation into *RHS*/pseudogenes, it is worth considering the different arrangements of retroelements that were found inserted in the *RHS* pseudogenes in the *T.brucei* genome. Inspection of the obtained sequences of *RHS* genes with insertions showed a number of overall possible conformations of insertion: Retroelements (ingi or RIMEs) could be inserted in different combinations: full ingi, truncated ingi, full RIME or truncated RIME (which may correspond to RIME A or RIME B) (See Figure 5-5).

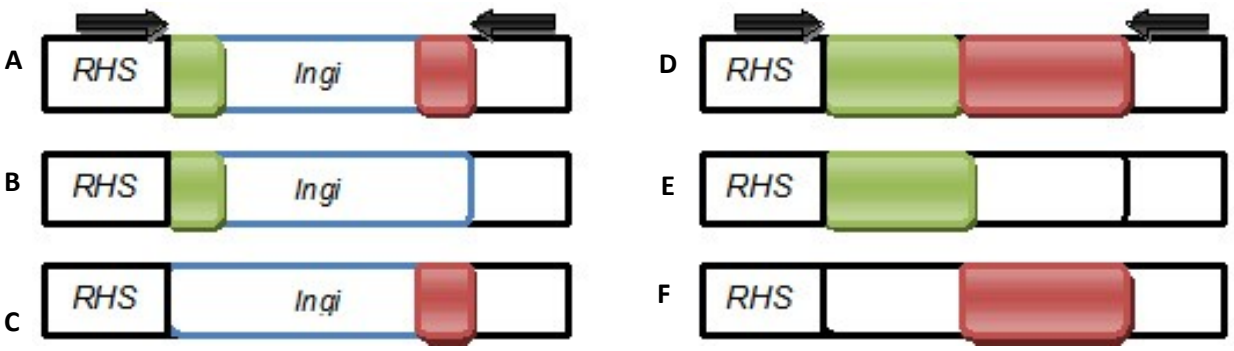


Figure 5-5. Schematic diagram of the RHS pseudogenes inserted with retroelements. (A) Indicates a full *ingi* insertion with the green and the red colours indicating the truncated ends of the *ingi* by RIME A and RIME B respectively. (B) shows the truncated *ingi* insertion where RIME B is missing, (C) illustrates the truncated *ingi* insertion where RIME A is missing, (D) shows the full RIME insertion without any component of *ingi*, (E) half RIME (RIME A) insertion, and (F) half RIME (RIME B) insertion. Black arrows represent the 12 bp repeat generated by the reteroelement insertion.

Most retroelements share several properties including (i) variability in copy number; (ii) insertion site polymorphism; and (iii) the duplication of a few base pairs of host DNA at each end of every new element insertion site (the hallmark of retroelement insertion) (Margaret *et al.,* 2001). *ingi* and RIME have been proven to generate repeat flanking in their insertion that varies from 8bp to 12bp (Bringaud *et al.,* 2004). This is important to highlight since some of the *RHS*/ pseudogenes do not possess a retroelement, but leave behind the flanking repeat when they are excised. These traces of the retroelement could help the investigate the evolutionary process of the *RHS* pseudogenes by indicating where an element has historically been situated. For example, the following sequence of events, *RHS* to *RHS* + Insertion (=pseudogene) to *RHS* + traces of insertion (=pseudogene), follows a temporal sequence that could be used to infer a timing sequence of evolutionary events for a given gene. In an epidemiological context, this could be used to track the evolution and transmission of field strains with time.

5.4.2.1 *RHS* 3 and *RHS*4 mutational change analysis

Mutations are the cause of the most genetic variation and cover changes in a single replicating genome, whether they are caused by DNA replication errors such as base pair substitutions, insertions and deletions or by the activity of transposons and insertion sequence elements that can move around (and replicate) in the genome.

5.4.2.1.1 Gene duplication

Gene duplication is an event in which one gene gives rise to two genes that cannot be operationally distinguished from each other (Roy and Deo, 2014). Gene duplication

can be classified into three categories; (i) one copy may simply become silenced by degenerative mutations (non-functional); (ii) one copy may acquire a novel, beneficial function and become preserved by natural selection with the other copy retaining the original function (neo-functionalisation); or (iii) both copies may become partially compromised by mutation accumulation to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (*sub*-functionalisation) (Lynch and Conery, 2000). One duplication of particular interest in the phylogenetic tree is in chromosome 4 of the TREU927 genome where a series of duplications took place (see Figure 5-6)

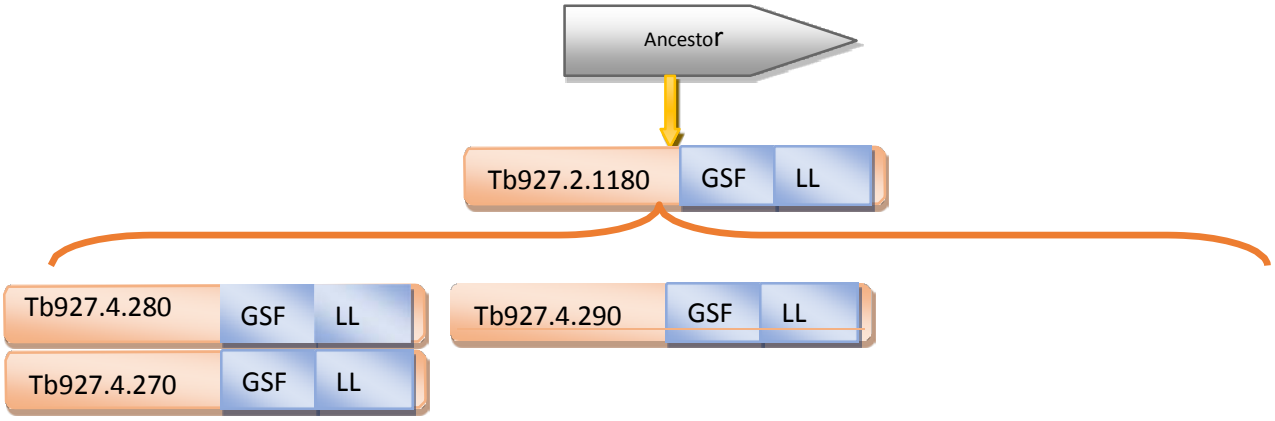


Figure 5-6. Diagrammatical representation of the consequences of duplicated genes and the role of duplication in the evolution of RHS genes using the RHS gene in chromosome 4 of TREU 927 genome. An example of sub-functionalisation is highlighted in blue where the amino acid sequence shown (GSFLL) represents the retrotransposon insertion site.

The above figure is an exemplary illustration of gene conversion duplication of *RHS* genes in the TREU927 genome; gene conversion is known as a transfer of sequence from one paralogue to the other. The full nucleotide alignment is listed in Appendix 3. In Figure 5-6 only 3 duplicate *RHS* genes were illustrated, however there are two more other duplicated genes Tb927.4.260 and Tb927.4.300. According to the nucleotide sequence alignments the duplicated genes have adopted mutations at the N-terminal,

which formed a chimeric gene that contains a 5' extremity of *RHS*1 pseudogene and *RHS* 3 at the c-terminal (3' extremity). Once gene duplicates have diverged sufficiently so that they differ in their functionality (or non-functionality), gene conversion events can become deleterious - for example, by introducing disrupting mutations from a pseudogene into its functional duplicate. A substantial proportion of disease alleles in Gaucher disease result from the introduction of mutations into the glucocerebrosidase gene from a tandemly repeated pseudogene (Tayebi *et al.*, 2003). The duplicated genes evidently exhibited the retrotransposon insertion site, however due to the amino acids' base changes the duplicated genes have lost the ATP/GTP motif. The selection on this family of genes was estimated (comparing the parent gene and the duplicated genes) by estimating the ratio of non-synonymous to synonymous substitution rates using the formula ω = (dN, amino acid changing)/(dS, amino acid retaining). With ω <0, ω =1 and ω>1 indicating purifying, neutral and adaptive evolution (Yang*et al.,* 2000). Nonsynonymous and synonymous substitution rates (denoted as Ka and Ks, respectively) were estimated using the KaKs Calculator (1.2) (Zhang *et al.,* 2006), with the method of model averaging YN. This produced a result which showed *neutral* evolution (ω =1), suggesting that early in their history many gene duplicates experience a phase of relaxed selection or even accelerated evolution at replacement sites (Lynch and Conery, 2000).

5.4.2.1.2 Role of Mobile elements insertion in *RHS* genes evolution

According to McClintock (1978) mobile elements have the ability to induce mutational changes in the genome through transposition, and McClintock reported a genome

restructuring process mediated by TE activity as an essential component of the hosts' response to stress.

5.4.2.1.2.1 Chromosomal translocation

Mobile elements insertion (transposition) can cause of number of chromosomal structural changes, such as translocation and inversion. As Lim (1988) has reported, such changes in *Drosophila melanogaster* are shown to be mediated by *hobo* transposable elements. Gray (2000) has suggested two possible mechanisms by which mobile element-associated chromosomal rearrangements can occur: (1) indirectly by homologous recombination or (2) directly by an alternative transposition process. As shown in Figure 5-7, the first scenario could be applied to Tb927.6.5170*RHS* pseudogenes.

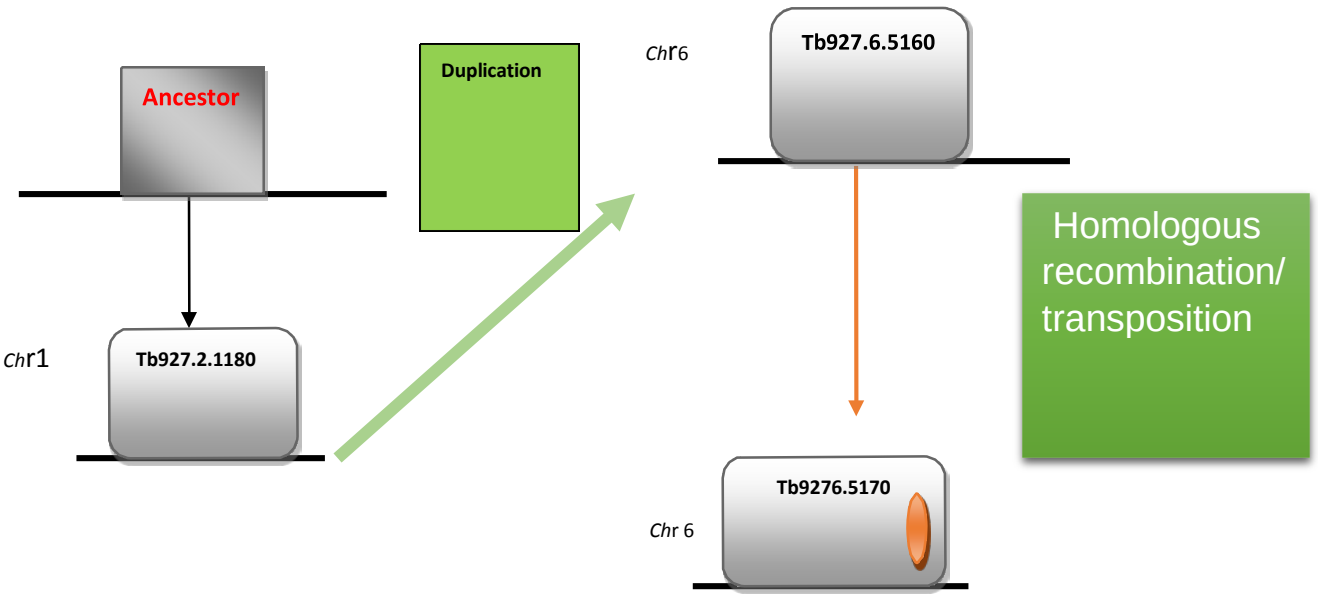


Figure 5-7. Schematic representation of putative RHS pseudogene duplication and chromosomal translocation of Tb927.2.1180, a functional RHS gene (Parent gene) in *T. brucei*. The duplicated pseudogene (Tb927.6.5160) is seen on chromosome 6 and the mobile element associated chromosomal translocation is shown in Tb927.6.5170 where the orange circle indicates a truncated *ingi* (RIME B is missing) insertion.

The first mechanism is more related to this type of duplication, whereby the indirect action of MGEs triggers chromosomal rearrangements by presenting the genome with multiple similar sequences between which strand transfers can occur. Moreover, multi-gene families possess a risk of homologous or ectopic recombination in which genetic exchange takes place, either intra-chromosomally or inter-chromosomally, between two members of the family located in different chromosomal regions (Kidwell, 2000). Such events can cause duplications or deficiencies.

On the other hand, Tb927.6.5160 is a pseudogene formed through the duplication of a functional gene (the parent gene being Tb927.2.1180) and a subsequent translocation from chromosome 1 to chromosome 6. Comparison of the DNA sequence of these two genes can be used to infer how recently this event has taken

place. For example, a recent duplication and translocation event would result in a very high sequence identity between both sequences whilst a less recent event would be shown by a greater divergence in DNA sequences. An alignment of both sequences is shown Figure 5-8. In this case sequence identity was determined as being at 88%. As Pei *et al.* (2012) have indicated, the higher the homology, the more recent the evolutionary event will have occurred.

```
Tb927.2.1180      ATGAATCAACAGGTTCCTATTGAAGGAAGGGGTGACATTGAGGGAAGGAGACGAGAAAAT
Tb927.6.5160      ATGAATCAACAGGTTCCTATTGAAGGAAGGGGTGACATTGAGGGAAGGAGACGAGAGCGT
                  ********************************************************    *

Tb927.2.1180      GAGGAGGCTGCGCGTAACGATGCGGAACCTCCAGTAGTGCAGCAACGAGTGGAGAATAAC
Tb927.6.5160      GAGGAGGCTGCGCGTAACGATGCGGAACCTCCAGCAGTGCAGCAACGAGTGGAGAATAAC
                  ********************************** ************************

Tb927.2.1180      AACCAACCACAATGGGGTCTTTTCAGCTGCGTTGACGCAGTTTTGCTGAATGGTCTGCCA
Tb927.6.5160      AAACAACCACAATGGACAATTTTGAGCAGCGTTGAAGCAGTTTTGCTGAATGGTCTGCCG
                  ** ************    **** *** ******* ***********************

Tb927.2.1180      CATCCTAGAAATATGATGTTGAATGATTTTTTGAGGCGTAATTTTGGCAGAAGATAT-AA
Tb927.6.5160      CATCCTGGAAATATGAGTGTAAATGATTTTTTGAGGTTTAATTT-ACTCGAAGATATCGA
                  ****** *********   * ***************  ******     ********  *

Tb927.2.1180      TGTGAATGAAGAAAAT---AATGTATCAATGTCTGCGTTCGTACTGGAACCTGAAGAGTA
Tb927.6.5160      TTTGGATCAGCGTCTCCTGAGAGCATCAATGTTTGCCTTTGTACAACGATGTGAAGAGTA
                  * ** ** *           *  * ******** *** ** ****    *  *********

Tb927.2.1180      TATTAATGACGTTAATGCGCTCAATCGAATATTCGCGACAACAGAGTATAAAGTGTACAA
Tb927.6.5160      TATTAATGACGTTAATACACTCAATCGAATATTCGCGGCAGTAGCGTATAAAGTGTACAA
                  ***************** * ****************** **  ** ***************

Tb927.2.1180      ACGTTTTGTCAGCGTCTATGGTTTCTTTGAGGATGAGGGAATATTGAATCTTCAGCGCTG
Tb927.6.5160      ACGTTTTGTCAGCATCTATGGTTTCTATGAGGATGAGGGAATATTGAATCTCAGGCGCTG
                  ************* ************ ************************   ******

Tb927.2.1180      GCAACAAGCAGATGAGGAAGCAAAGGTCAGGTTACAGGCGGATATTCGTGGACTGAGAGA
Tb927.6.5160      GCAACAAGCAAATGAGGAAGTAAAGGCCAGGTTACACGTGAATATTCGTGGACTGAGAAA
                  ********** ********* ***** ********* * * ***************** *

Tb927.2.1180      TGGAGAGCGCCTTTGGATTGTTGTCACTAACATGTTGAATGATGCACTCAATGAAGCAAG
Tb927.6.5160      TGGAGATCAAACTTGGATTATTGTCACTAACATGTTGAATGATGCACTCAAAGAAGCATT
                  ****** *    ******* ******************************** ******

Tb927.2.1180      GGAACGAGCTGCTCAAACAGCTTGTAGTGCTGTCGAACTGAAAGGATTGTATGAATCCAT
Tb927.6.5160      GGGAGAAGCTGCCCAAACAGCTGGTGGTGCTGTCGAACTCTCAGGTTTGTATGATTCCAT
                  ** *  ****** ********* ** **************   *** ******** *****

Tb927.2.1180      ATATAATGCGAAATGGAGTTATGTGATGTCGGGTTATCATACAGAGCCACTTGGCATGAA
Tb927.6.5160      TTATAATGCGAAATGGAGTTACGTGATGTCGGGTTATGATGCAGGGCCATTTGGAATGTA
                   ******************** *************** ** *** **** **** *** *

Tb927.2.1180      AGTGTTCGATGGAAGGCCGCAGCGCATGTGGACCGAAGAAGATGTGGATATAACTCCTCT
Tb927.6.5160      TGTATTCAATGGAAGACCACAGCGCATATGGACCGAAGCGGAGGTGGATATAACTCCTGA
                   ** *** ******* ** ******** **********  ** ***************

Tb927.2.1180      TCCTGCGAATGTTGATGCAGAGATTGAGGAGAGACCCGATGGTTTGGAGATTTTTGTTCT
Tb927.6.5160      TCCTGCGAATGTTGATGCAGAGATTGAGGAGAGACCTTATGGTTTGGAGATTTTCGTTCT
                  ***********************************  ***************** *****

Tb927.2.1180      CACTTCGGAGAAAGGTTGGCCATACAACAGGTTTGCATTGGATTACACTACGGGGCGCAA
Tb927.6.5160      CACGTCTGAGGAAGGTTGGCCATACAACAGGTTTGAAATGGATTACACTACGGGGCGCAA
                  *** ** *** ************************ * ***********************

Tb927.2.1180      AGTAGTATTTCAGCACGTATACATCCGTCGTGAAATTGTGCGCGTGTGGTATAAAGTCGA
Tb927.6.5160      AGCAGCATTTCAGCATGTATACATCCGTCGTGAAATTATGCGCGTGTGGTATATAATTCA
                  ** ** ********* ********************** *************** * *  *

Tb927.2.1180      AAAAGATCTGAAAACATGGTGGGTGGAAAAGACAGCTCACAGACCACCGATACATATTGT
Tb927.6.5160      ACGAGGATTGCAAACATGGTGGGTGGAAAAGACGGCTCACAGACCACCGATACATATTGT
                  *  **   ** ********************** ***************************

Tb927.2.1180      TATTGGTACACCTGGTATTGGTAAATCATATGGCCTTGGATCATTTTTGCTTCATTCGTT
Tb927.6.5160      TATTGGTACAGCTGGTATTGGTAAATCATGTGGTGTTGGATCATTTTTGCTTCATTCCCT
                  ********** ****************** ***  ********************** *

Tb927.2.1180      GCTTCACTTCCATGAAGGAATGCTTGATGTTGTTGCATATTTCACAGACACGATCGCCTA
Tb927.6.5160      GCTTCACTTCCATGAAGGAATGCTTGATGTTGTTGCGTATTTCACAGACACGATCGCCTT
                  ************************************ **********************

Tb927.2.1180      CTTAATATATAACAGAAAGGGTGATGAAAGAGGGAGGGTTGTGCGGTACGAATATTTGAG
Tb927.6.5160      CCTCATATACAACAGAAAGGGTGATGAAAGAGGGAGGGTTGAGCGGTACGAAGGTTCGAG
                  * * ***** ******************************** ********** ** ***

Tb927.2.1180      AGCTGCGGTCAACGCCATAAATAAAATGAAATTTGAAAACAGAGGACACATTATTATGGA
```

```
Tb927.6.5160      AGCTGCGGTCAACGCCATGAATAAAATGAAATTTGAAAACGGAGGTCACATCATTATGGA
                  ****************** ******************** **** ***** ********

Tb927.2.1180      CATAAGGTATGCAATGCAGCAGCTTTACACTCAACTTCCCTCTGATGTTTGGAGTGTAAC
Tb927.6.5160      CACAAGGTACGCAACGCAGCAGCTGTATACTCATCTTCCCTCTTATACTTGGGGTGTAAC
                  ** ****** **** ********* ** ***** ********* **  **** *******

Tb927.2.1180      TATTTTCACTTCCCCGAATAGTTCCCACTTTGGTAACTGGACTACAATTACGGGAGGCCG
Tb927.6.5160      TTTCCTCACTTTTCCGAATAGTGCCCGTTTTGATGACTGGACTACAAATAGGGGAGGCCG
                  * *  ******  ********* ***  **** * ************ ** *********

Tb927.2.1180      CCAAATTATTATAAACTGTGATGATGTGCGTGACATGAAGGCATTTGTGGCATGGAAAAA
Tb927.6.5160      CCAAATTATTATAAACTGTGATGATGTGCGTGACATGAAGGCATTTGTGGCATGGAAAAA
                  ************************************************************

Tb927.2.1180      ACTGTCAATACATACAAGGGAAAAGGTATCCAACAGAAGAAGACACGAATTGAGGAAGGA
Tb927.6.5160      GCTGTCAATACATGCAAGGAAAAAGGTATCCAACAGAAGAAGACACGAATTGAGGAAGGA
                   ************ ***** ****************************************

Tb927.2.1180      AATGGAAGATGAGTGGAGAATAGTGGAGGGACGTATTAATTCAATTGGGCTTTTGCCTCG
Tb927.6.5160      AATGGAAGATGAATGAAAAATAGTGGAGGGACGTATTAATTCAATTGGGCCTCTGCATCC
                  ************ ** * ******************************** * *** **

Tb927.2.1180      TTATATATTTGTTTTAAGCTGTTATGAATGGCGTCTGAAAAGAGTTCATGATGCGTTGGA
Tb927.6.5160      TTATATATCTGGTTTAAGATGTTATGAGTGGCGTCTGAAAAGAGTGCATGATGCGTTGGA
                  ******** ** ****** ******** ***************** **************

Tb927.2.1180      AACTATGAAGAAGTCGGATGAATATTCCTATAATGATATTATTGAGCATACCGCTGCCTG
Tb927.6.5160      AACTATGAAGAAATCGGGTGAATATTCCTATAATAATATTATTGAGCATACGGCTGCCAG
                  ************ **** ***************** ***************** ****** *

Tb927.2.1180      GAAGAACAATGAAGTTACGGAAAAGTTGGTAAAGGTTGTAAGAGTAAAGGGAAATGTTGG
Tb927.6.5160      AGAGAACAATGAAGTTATGAACAAGTTGGAGAAGATTGTAAGAGTAAAGGCAAATGTTGG
                    *************** * * *******  *** *************** *********

Tb927.2.1180      ATTTATCGAATCATTTAAATTTCAAGCTCTATCACTGATGATTCGAAATATGATGATGAG
Tb927.6.5160      AGGAATCGAATCATACAAATGTCAAGCTCTATCACCGATGATCCGACATATGATGATGAG
                  *   **********  **** ************** ****** *** **************

Tb927.2.1180      TTAA
Tb927.6.5160      CTAA
                   ***
```

Figure 5-8. DNA sequence alignment between the parent gene (Tb927.2.1180) and duplicated pseudogene (Tb927.6.5160). The alignment was carried out using EBI clustalW2 with default settings. The identity is calculated at ~88%. Base pair identities are shown by*.

5.4.3 Evolutionary time dimensions of *RHS* genes

The transposition (insertion) of non-LTR retrotransposons, including *ingi*s or RIMEs, produces a duplication of the target sequence to form a direct repeat of a few bases (11-14bp) flanking the inserted retroelement. In terms of *T. brucei*, Bringaud *et al*. (2002) have reported that the formation of 12-bp sequences are possibly the result of multiple *ingi/*RIME insertions at the same site with the consequence being multiple duplications of the target site, each flanked by retroelements. Based on these facts we hypothesised that we could categorise the evolution of *RHS*/pseudogenes according to where the mobile element was inserted and excised and when a fingerprint was left with the 12bp flanking sequences as evidence for their past-existence (Figure 5-9).

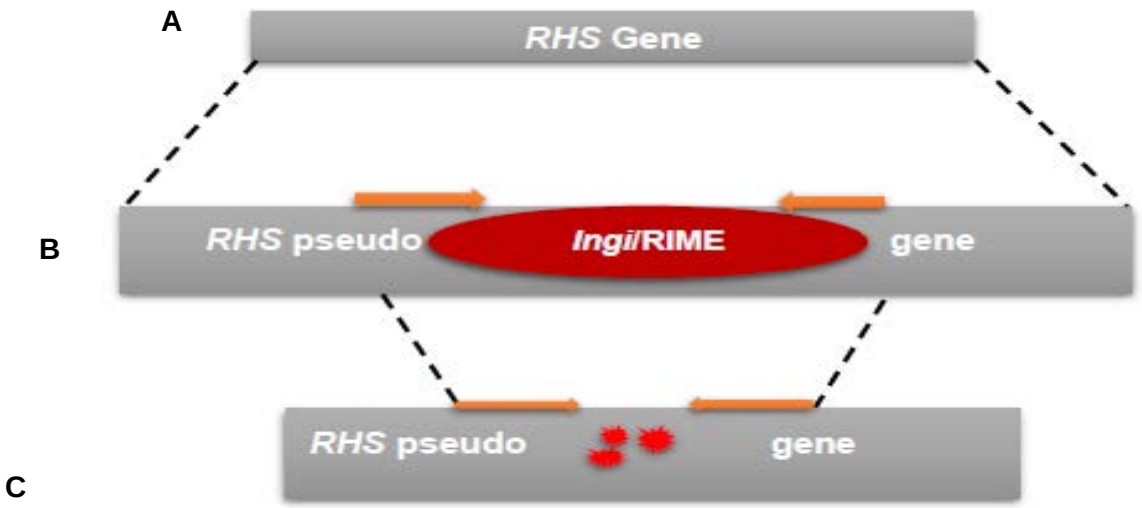


Figure 5-9. Schematic diagram illustrating a possible evolutionary scenario for the RHS pseudogenes. (A) is an RHS gene, (B) is a RHS pseudogene that possess a mobile element insertion highlighted by the red oval (this could be either ingi or RIME), and (C) a RHS pseudogene that possess the flanking direct repeat highlighted by two opposite arrows and fingerprints of the mobile element traces highlighted by red (explosion marks). Dashed lines indicate the sequence similarities

According the above figure which shows possible categorisation based on an evolutionary time dimension, *RHS* genes/pseudogenes can be classified into three classes: (1) original state (A in Figure 5-9), (2) intermediate state, the transposition of the

retroelement has occurred (B in Figure 5-9) and (3) recent state whereby the MGEs have been excised and left fingerprint traces or did not leave any traces in the *RHS* pseudogene (C in Figure 5-9). Therefore, MGEs' insertion in *RHS*3 and *RHS*4 were screened for the duplicated repeat sequences flanking the inserted MGEs (Table 5-1)

Gene Name	5'	Insertion target	MGE type	Duplicated Sequence
Tb927.4.220	AGTTGGTGTTGGCTCGT	TCCTGCTCCAA	RIME	TCCTGCTCCAA
Tb927.5.280	TGAGAGCATGGTGGGTA		Truncated *ingi*	*ingi*
Tb927.6.5170	TCATTTTTGCTTCATTC		Truncated *ingi*	
Tb927.6.5180	Truncated *ingi*	TTTTGCTTCATT	RIME	TTTTGCTTCATT

Table 5-1. The duplicated repeat sequence flanking the MGE insertions in RHS pseudogenes in the TREU927 genome.

As can be seen from Table 5-1, the duplicated sequences are flanking the MGEs insertion, and Bringaud *et al.* (2004) have suggested that this 12bp repeat is the result of a multiple *ingi*/RIME insertion, which is related to Tb927.5.5180 where multiple insertions have occurred, whereas Tb927.5.280 and Tb927.6.5170 do not possess a duplicated flanking repeat, which is in accordance with the work of Bringaud *et al.* (2004) wherein the authors reported that not all the MGEs are inserted between two flanking repeats.

The hypothesis was thus investigated using three methods: first by aligning each *RHS* pseudogene in Table 5-1 with its sequence related *RHS* gene in accordance with the phylogenetic tree (Figure 5-4); secondly a further alignment was carried out between each of the *RHS* pseudogenes in Table 1 and its related sub-family, for example Tb927.6.5180 was aligned with all the *RHS*4 genes in the TREU927 genome; thirdly and finally a BLAST search with the insertion target nucleotide sequences of *RHS*3 and *RHS*4 was carried out.

The result of the final experiment did support the original hypothesis, as the recent state of the evolutionary *RHS* was found in *RHS*4 subfamilies where the two sets of

the flanking direct repeat were present showing that the MGE had been excised, as it has shown that the Tb927.1.70 (*RHS* Gene), that possesses the duplicated insertion target (12bp direct repeat) at both the 5' and 3' end, which suggest that the excision of RIME has occurred, through possible homologous recombination (Bringaud *et al.,* 2004).

An alignment between Tb927.6.5180 and Tb927.2.1210 (first method) showed that Tb927.2.1210 possesses a MGEs insertion target, and it was shown that Tb927.2.1210 has the first set of the 12bp flanking sequences. However, the presence of the same sequence of the insertion target (12bp repeat) in the Tb927.2.1210, could be considered as evidence for a past insertion and excision (Figure 10). In order to discern whether the traces of MGEs were present in Tb927.2.1210, further alignment between Tb927.2.1210 and the MGEs (*ingi* and RIME) was carried out. The results did not confirm the presence of traces.

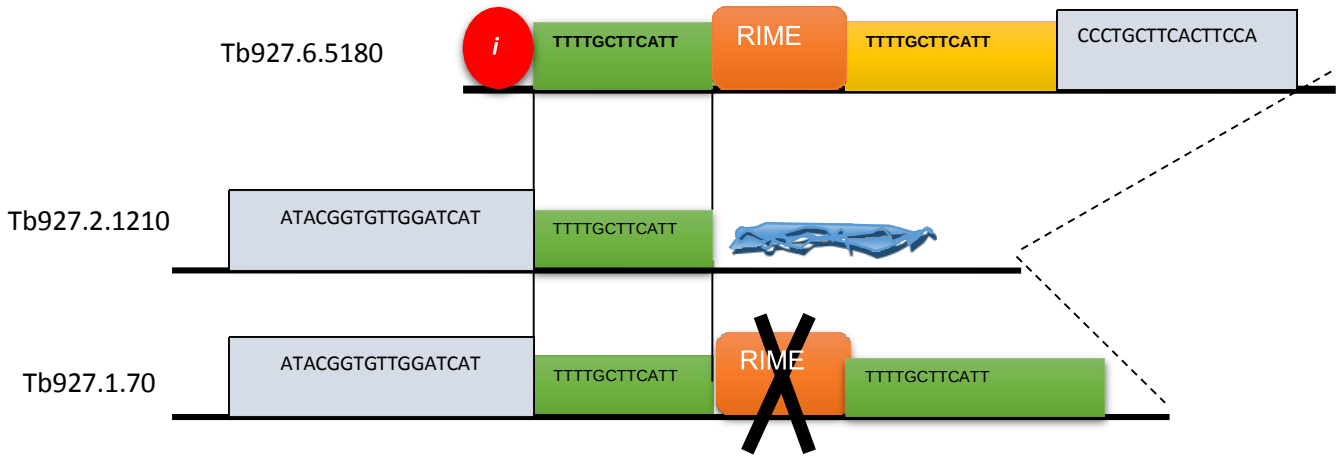


Figure 5-10. Schematic diagram of the duplicated 12bp repeat similarities between Tb927.2.5180, Tb927.2.1210 and Tb927.1.70. The straight line represents the absolute alignment and the dashed line indicates high identity between the three sequences. and *i* represents the truncated *ingi*. Crosses indicate the excision of RIME, and the blue scribble indicates the sequence which is not related to any of the MGEs. The duplicated repeats (5' and 3' ends) that flank the MGE (insertion target) are in green boxes. The *RHS*/pseudo genes nucleotide sequences are indicted in blue boxes.

In summary, the amino acid sequence alignment of the *RHS* pseudo genes (Appendix 2) illustrated that the majority of the *RHS* genes have a retroelement insertion target site. However, the presence of one set of the duplicated repeat has resulted from the insertion of MGEs and can be considered as evidence of MGE excision, and it can be concluded that Tb927.1.70 is in the "recent state" in terms of the evolutionary time dimension of the *RHS*/pseudogenes in the *RHS*3 and *RHS* 4 subfamilies

.

5.4.4 Mobile genetic element markers in the differentiation of trypanosome strains

Mobile genetic elements have been proven to be reliable genetic markers. Thus, a possible MGE marker that can differentiate between *T.b.brucei* and *T.b.gambiense* was investigated based on the method of retro-insertion analysis. Retro-insertion is defined as a comparative investigation of the presence or absence of retroelements at multiple loci in the genomes

(Shedlock *et al*., 2004). According to Jackson *et al*. (2010) intraspecific genome divergence between *T.b.brucei* and *T.b.gambiense* is estimated to be 1% where the indels or insertions are generally caused by mobile elements.

Once such insertion sites have been identified, sequences flanking the MGEs at a specific locus can be used as primers to amplify genomic DNA isolated from a parasite. When the PCR products are visualised an agarose gel, individuals with insertions at both alleles will show a single high molecular mass band while individuals with no insertions at either allele will give a single low molecular mass band (Tu and Li, 2000). *Alu* elements have been shown to be extremely powerful genetic markers using this locus- specific PCR method (Stoneking *et al.,* 1997). In this section retro-insertion will be studied in *T. b. brucei* and *T. b. gambiense*, using BLAST searches on the NCBI database. A comparison will be made between all the *RHS* pseudogenes that are inserted with a MGE present in the *T. b. brucei* TREU927 strain (see Table 2) and the correspondent *RHS* pseudogenes in the *T.b.gambiense* genome (DAL 972). The sequences of the flanking region of the *RHS* pseudogenes will be examined for sections that are identical or which show high similarity between the subspecies. Candidate genes for further analysis will be selected from those where the MGEs are missing in the DAL 972 gene compared to the corresponding TREU927 gene. This can help to find a potential marker which could be used in differentiating between *T.b.brucei* and *T.b.gambiense.*

RHS Gene	Location	RIME	*ingi*
Tb927.2.1150	Chromosome 2	RIME A	
Tb927.2.1160	Chromosome 2	RIME B	
Tb927.4.100	Chromosome 4	RIMEA	
Tb927.4.110	Chromosome 4	RIMEB	
Tb927.4.210	Chromosome 4	RIMEA	
Tb927.4.220	Chromosome 4	RIME B	
Tb927.5.280	Chromosome 5		Truncated ingi
Tb927.6.5170	Chromosome 6		Truncated ingi
Tb927.6.5180	Chromosome 6	RIME	Truncated ingi
Tb927.10.110	Chromosome 10	RIME	
Tb927.11.4770	Chromosome 11		Truncated ingi

Table 5-2. The RHS genes that have an MGE insertion.

Out of the 59 *RHS* genes only 11 have the *ingi*/RIME insertion Tb927.11.4770 has the truncated *ingi* insertion and another mobile element known as *DIRE* small degenerate retroposons (0.55 kb) that have been found in genomes of *L.major* and named LmSIDERs. The results of the BLAST searches revealed, that out of the eleven *RHS* pseudogenes, 4 *RHS* pseudogenes from TREU 927 (Tb927.5.280, Tb927.5.5170, Tb927.6.5180, and Tb927.11.4770) could be considered as potential markers, since their corresponding *RHS* pseudogenes in DAL 972 (*T.b.gambiense*) lacked a MGE insertion. However, due to the relatively small size of the RIME, the presence of the hallmark of retroelements in the inserted RIME, and the presence of the duplicated 12bp repeats in Tb927.6.5180, this gene is the most suitable *RHS* pseudogene as a potential marker. There was a high homology shown by the *RHS* pseudogenes flanking sequences of Tb927.6.5180 and Tbg972.6. 4970 (see Figure 5-11).

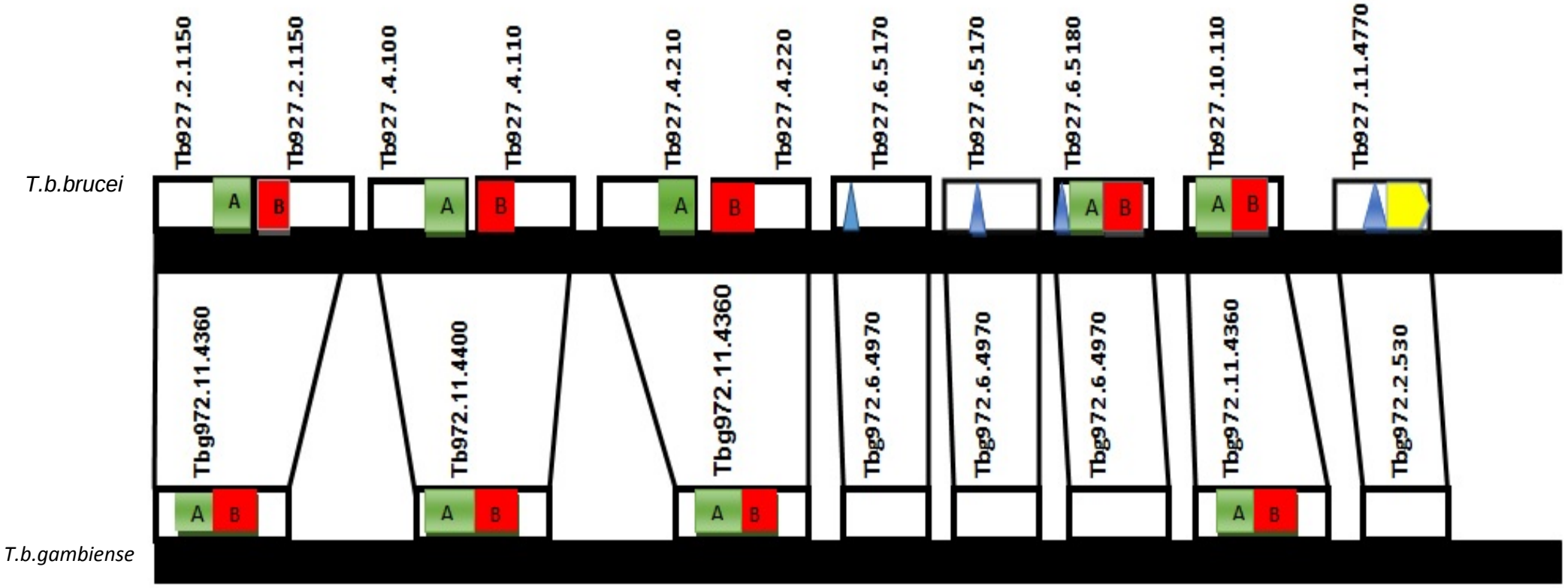


Figure 5-11. Schematic diagram of the nucleotide comparison between RHS pseudogenes of the TREU927 (upper line) and DAL972 (lower line) genomes. The MGE insertions are indicated by (A), (B), blue triangle, and yellow pentagon, representing RIME A, RIME B, truncated *ingi* and *DIRE* respectively. The straight lines indicate the high homology alignments. Where the A or the B is present on their own, this indicates half of a RIME insertion.

The purpose of this comparison was to identify sequences that could be used to generate PCR primers from the RIME and its flanking DNA sequences to distinguish between the two subspecies. In greater detail, the aim was to identify a forward primer from the Tb927.6.5180 flanking sequences and a reverse primer from the homologous flanking regions of Tb927.6.5180 and Tbg972.6.4970. The size of the amplicons would be different due to the insertion of the RIME element in the *T. b. brucei* sequence. Figure 5-12 shows a CLUSTALW alignment of the two sequences to demonstrate the regions of homology and difference.


Tb927.6.5180 CAACGCGGCGACTG--CGAAGAAACGTCGCTGCACTTGATGAGCTGTCCCGGTTTGAAGG
Tbg972.6.4970 ------GATTATTATTCGGAGAG-------TGAAATGAATCAACAGGTCCCTATTGAAGG
Tb927.6.5180
AGTTACGTGTCAGATTTGCGGTAGAAGGTGAGTGTGTGCAGGATGTATGCTTCTCCAAAA Tbg972.6.4970
A---AGGGGTGACAATGAGGGAAGGAGACGAGAGCGTG-AGGAGG----CTGCGCGTAAC
Tb927.6.5180 GGCTGGCGCAGTTTCT--CATTGCGGT--TGAACGGAGC--------CGGCCGCAGGTGA
Tbg972.6.4970 GAT--GCGGAACCTCCAACAGTGCAGCAACGAGTGGAGAATAACAAGCAACCACAATGGA
Tb927.6.5180
-AGTCCTCACCTAAGGTAA--CAGTCATACAACCCACTCTCCC---TTCTGCAACTTCCC Tbg972.6.4970
CAATTTTGAGCAGCGTTGAAGCAGTTTTGCTGAATGGTCTGCCGCATCCTGGAAATATGA
Tb927.6.5180 CCCTTATTG-----CCGAGTGTGGAGCAAGCAGGAAAAGCACCGGACGCAGGAATAGTCC
Tbg972.6.4970 GTGTAAATGATTTTTTGAG-GTTTAATTTACTCGAAGA-TATCGATT-------TGGATC
Tb927.6.5180 AGACACCCTCAGAGACAGAGGAGGTATGCAGTCA-ACAAGCAACAGAAACAGAAAGAGGG
Tbg972.6.4970 AG-CGTCTCCTGAGAGCATCAATGTTTGCCTTTGTACAA-CGATGTGAAGAGTATATTAA
Tb927.6.5180 GGA-GAGAATAA---TAAACGAATA---ACAAAAATCAAAACAAAAGGATTGCTAA----
Tbg972.6.4970 TGACGTTAATACGCTCAATCGAATATTCGCGACAAT-AGCGTATAAAGTGTACAAACGTT
Tb927.6.5180 TTGACATCTTTGGGAGAGTCCGGGGTGAGGGGGGCTTCTCG--CCCCATCTGCTGTATTC
Tbg972.6.4970 TTGTCAGCATCTATGGTTTCTATGAGGATGAGGGAATATTGAATCTCAGGCGCTG-GCAA
Tb927.6.5180 CGTTCAACTGCGGAGCTACAACAAAAATTATAGAGGGTGTGTTAGGATGAATAAAAAAGG
Tbg972.6.4970 CAAGCAAATGAGGAAGTAAAGACCAGGTTACAC-GTGAATATTCGTG-GACTGAGAAATG
Tb927.6.5180 GAGACTCTGCCACAGTCGCCAGACCGATAGCATCTCAGGGCTCTACGGTGATG-GCTGAT
Tbg972.6.4970 GAGATCAAACTTGGATTATT--GTCACTAACATGTTGAA---------TGATGCACTCAA
Tb927.6.5180 GGCCGCGCCAGTTGGGAGAAACTCTCACGAAGGCACGAAGAAAATTCAAAAAAAAAAATT
Tbg972.6.4970 AGAAGCA----TTGGGAGAAGCTGCC-CAAACAGCTGGTGGTGCTGTCGAACTCTCAGGT
Tb927.6.5180 TTGCTTCATT[redacted]CCGACCACCTCAACGTGGTGCCAGGGTCCAGTACCCCG
Tbg972.6.4970 TTGTATGATTCCATTTATAATGCGAAATGGAGTTACGTGATGTCGGGTTATGATGCA--G
Tb927.6.5180 TATCATCGGGG------GAAGCCAA---GAGCCAGCAGCGTTCCTTTCATGGGGAACACT
Tbg972.6.4970 AGCCATTTGGAATGTATGTATTCAATGGAAGACCACAGCG-------CATATGGACCGAA
Tb927.6.5180 GCTGTGCTCCGGCTACGGCATC--ATACAGCACAGG--GATCAGCAGCGTCTTGCTGGGA
Tbg972.6.4970 GCGGAGGT--GGATATAACTCCTGATCCTGCGAATGTTGAT--GCAGAGA-TTGAGGAGA
Tb927.6.5180 CACCGTTTTTCATTTGTCGGTCCCTGGGCACGTGCCAGCGTGCCATCAGCAGTATCATCC
Tbg972.6.4970 GACC--TTATGGTTTGGAGATTTTCGTTCTCACTTCGGAGAAGGGTTGGCCATA-CAACA
Tb927.6.5180 GCACTAAGATG--CTGCTGTCCGGTGATGTGGACCTCCAAAAAAGGATTGCCAATTGGCA
Tbg972.6.4970 GGTTTGAAATGGATTACACTACGG-GATG--------CAAAGCAGCATT-----TCAGCA
Tb927.6.5180 TCTTTTGGAGAGTCCAGGGTGGGAGGCTTCTCGCCCCATCTGCTGTATTCCGTTCATATG
Tbg972.6.4970 TGTAT----ACATCC--GTCGTGAAATTATGCG---CGTGTGGTATAT------------
Tb927.6.5180 CGGAAATACAACAAAAATTATAGAGGGTGTGTTAGGATGAATGAAAAAGGGAGACTCTGC
Tbg972.6.4970 ----AATTCAACGAGGGTTGCAAA--------CATGGTGGGTGGAAAAGACAG-CTC---

```
Tb927.6.5180       CACAGTCGCCAGACCGATAGCATCTCAGGGCTCTACGGTGATGGCTGATGGCCGCGCCAG
Tbg972.6.4970      -ACAG--ACC--ACCGATA-CAT----------------ATTGTTATTGGTACAGCTGG
                    ****   **  ******* ***                ** * *  ***    **  *

Tb927.6.5180       TGGGGGGAAACTCTCACGAAGGCACGAAGAAAATTCTAAAAAAAAAAATTTTGCTTCATT
Tbg972.6.4970      TATTGGTAAA---TCATGTGGTGTTGGA------TC---------ATTTTTGCTTCATT
                   *   ** ***   *** *  *    * *      **         * ************

Tb927.6.5180       CCCTGCTTCACTTCCATGAAGGAATGCTTGATGTTGTTGCGTATTTCACGGGCGGGATAG
Tbg972.6.4970      CCCTGCTTCACTTCCATGAAGGAATGCTTGATGTTGTTGCGTATTTCACAGGCGGGATAG
                   ************************************************** **********

Tb927.6.5180       CCTACTTAATATATAACGATAAACCCGGTGAAGAGGGGAGAGTTGAGCAGTATAAAGGCG
Tbg972.6.4970      CCTACTTAATATATAATGATAAACCCGGTGAAGTGGGGAGGGTTGAGCAGTATAAAGGCG
                   *************** *************** ****** ******************

Tb927.6.5180       TGACAGCTTCTGTGAGGGGAATAAATGCAGTAAAGTACGAAAAGACAGGTCACATCATTA
Tbg972.6.4970      TGACAGCTTCTGTGAGGGCAGTAAATGCAGTAAAGTACGAAAAGACAGGTCACATCATTA
                   ***************** * ************************************

Tb927.6.5180       TGGACACAAGGTACGCAACGCAGCAGCTGTATACTCATCTTCCCTCTTATACTTGGGGTG
Tbg972.6.4970      TGGACACAAGGTACGCAACGCAGCAGCTGTATACTCATCTTCCCTCTTATACTTGGGGTG
                   ************************************************************

Tb927.6.5180       TAACTTTCCTCACTTTTCCGAATAGTGCCCGTTTTGATGACTGGACTACAAATAGGGGAG
Tbg972.6.4970      TAACTTTCCTCACTTTTCCGAATAGTGCCCGTTTTGATGACTGGACTACAAATAGGGGAG
                   ************************************************************

Tb927.6.5180       GCCGCCAAATTATTATAAACTGTGATGATGTGCGTGACATGAAGGCATTTGTGGTATGGA
Tbg972.6.4970      GCCGCCAAATTATTATAAACTGTGATGATGTGCGTGACATGAAGGCATTTGTGGCATGGA
                   ************************************************** ******

Tb927.6.5180       AAAAGCTGTCTGTAGCATCATGCTATGCGATACTTGATAAAGAGTCACAGGTGAGGATTG
Tbg972.6.4970      AAAAGCTGTCTGTAGCATCATGCTATGCGATACTTGATAAAGAGTCACAGGTGAGGATTG
                   ************************************************************

Tb927.6.5180       TAGAAGAAATCAAAAATGAGTGGAAAACAATAGAGGGCCGTGTTGATGTAGTTGGGCCAG
Tbg972.6.4970      TAGAAGAAATCAAAAATGAGTGGAAAACAATAGAGGGCCGTGTTGATGTAGTTGGGCCAG
                   ************************************************************

Tb927.6.5180       TGCCTCGTTGCGTTTACGACAGCATGGAGTACACAAAACGTGTGTCTCAGGCAAAAAAGA
Tbg972.6.4970      TGCCTCGTTGCGTTTACGACAGCATGGAGTACACAAAACGTGTGTCTCAGGCAAAAAAGA
                   ************************************************************

Tb927.6.5180       CGATAACGGCCATCAATGATGATAACAAAAGACACTACGAGGACATAATGATGAAAATTG
Tbg972.6.4970      CGATAACGGCCATCAATGATGATAACAAAAGACACTACGAGGACATAACGATGAAAATTG
                   ************************************************ ************

Tb927.6.5180       CCGGTTGGCAAAATGATAAGGTAACGCACAGGTTAGTGAGAATCGTTCGTGTGCGAGATG
Tbg972.6.4970      CCGGTTGGCAAAATGATAAGGTAACGCACAGGTTAGTGAGAATCGTTCGTGTGCGAGATG
                   ************************************************************

Tb927.6.5180       CCGGAGAGAGCATCGATGGATACCGCTGCAGGCCACTTTCGGCTTGCATTGGGAA-TGCT
Tbg972.6.4970      CCGGAAAAAGCATCGATGGATACCGCTGCAGGCCACTTCCGGCTTGCATTGGGAAATGCT
                   ***** * ****************************** *************** ****

Tb927.6.5180       ATATTTTTAAAACTGCTTTCCATTATTGTGCAGAAAATGGCTATCCGTCAAGTTATGATT
Tbg972.6.4970      ACATTTTTAAAACTGCTTTCCATTATTGTGCAGAAAACGGCTATCCGTCAAGTTACGATT
                   * ********************************** ***************** ****

Tb927.6.5180       TCAGACGAGAGTGCTGCGGCCAACGTTTTCAAATCGGGGGCGTTGTTTGCATTTCTTTTT
Tbg972.6.4970      TCAGACGAGAGTGCTGCGGCCAACGTTTTCAAATCAGGGGCGTTGTTTGCATTTCTTTTT
                   ********************************** ************************

Tb927.6.5180       CCTCGAGTTTTTAAAGTTATCAGAGACAACATAAAATACCTGCGGCGGCCGGGTGGACCG
Tbg972.6.4970      CCTCGAGTTTT-------------------------------------------------
                   ***********
```

```
Tb927.6.5180       GAAGAGAAACGATGCATACTGAAAGACATGACTCCTCAGCAGTTGAGACTCACTGGGCAG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AAGTTTTTACCAGATGCCAGACAGCAGCCAATGGAGAAGTGTGAATACATGGTTCTCTAC
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CGCCCTGCCACAGTCAATGAACCCGTGGTGGATGGCTTTTTTTTCGTTGAGGGGCGGCTC
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AGGAAAACTCCTAAGGGCAGGGCGGCACTGACCACTCCAAATATAGCTGTGCTTTTACAA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       GTAACCAAGTTGGCATGTCACCCCACCACTGCGTCAAAGGTGCAGAAGTTCCGAGAAAAC
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       ATGGCAAGATATTTTTCTGATTGGGGTGCGTTTTCCAGCAACATGGTTTGGGAGATGATA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TACATCAATGGAGCTAGTGGTGGTGTGATTACAAGATGGCAGCTTTGTGATGGCGACGAC
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CCCGCCGAAGCGGCTGATATGTGGAGAAACATGACTCAATACCAAGTAACGTTAAGAGAA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       GAAATGCAGGGGCAGCTAATGCGAGCATATGCAGAAGAAGGTTGGTATCGGGATGCTCCA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CTTCTGGCCATGGAGCCTGCGCCGCAGCCACGCAACGAGCAAGCGGGAGCGCAAGGAGGG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TGAGGCGATTTCACAGCATTGAGTGAGGTAGAAGCTGAGTAAAAATATCCGGCCTGGTAT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       GTTATGGTGGACGCGAGCAGTCAACAGTGTACGGTGCTAATGTGGGGCGAGGCGGCGTTT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TTTCCATTCAAAGTCAAATTGAGTTATCGTATCATTGGGAGGCCAGCAGTCCTAACTGCC
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AAAACAGGGTAAATGGCTTTGTCTTAATGCTTGACATGACAGGATGCGCTGCAAGGGAGT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AATATTCAGCAAATAGCGGCTATAGTAATCATCTTATTCTTTTCAACAAGCAGTCAAATG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CGGCGGTTAGTGCAGGAAAGGGACGTGATAATTCGCTTGAGTGTCGGGAGGGATGGGAGG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       GAAGTGGAAATGAAGACGACTAAGGAATAGTAAAAGTAGTATTTTATCGGCTAACGCGTA
Tbg972.6.4970      ------------------------------------------------------------
```

```
Tb927.6.5180       ACTGGAGTTAAGCCACATCGTAATCTTAAGGGGCACAGAGAAAGACAGAAATAATGTTAT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CGACGGACGTAAGAAAACATTGTAACGTTTTGTTTTCATTTTTGCTATTACCAACTCAAG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AATTGGATTATACTAGACTAACTCTTCACATCTTCCTTTCATTTATACAATCTAAATTTT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CCATCCGTGTAGAGATAGGAATCTATAGTCCAATGATAAAAGGAACAAAAATGTAATAGT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       ACTTTTCGCTTCTTCGCCATCCGTGGTGGAGAGAGAGTATTGGAAACATGCCTCAACTCT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       AACTCTGCTGAATCCTATTTCCAGTAGGTGTAATGGGAGAGGGATTGTGACGGAGTGAAA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TATGCAAACATTAGAAGCAGTTACTGGATGTGCTGTTTCATATACGAGCTGTTATAATTA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TATTATACATGTTAGTTATGGTTGCAGTTATGGAGCAGCTCTTGATGGTGTTCTTCGCAG
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       ACGCGCAGCTGGTGACGTGATTGTGTCCATTTTCCTCTCTCACTCTCTTTCTTTCTTCAA
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TTATTTGTCGGTGTGCAAACATACTATGATATGTTGTCTTCTCTTTACTTATAATCACTT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       CTGTTATCCTGGTTGAGGGTTGGCAGTTGGGCCTCTCTCCACACGTGGCAACCTTAAACT
Tbg972.6.4970      ------------------------------------------------------------

Tb927.6.5180       TATGTTGCTGGTTGCCTCCAGCAGCAAATCCCATATCATCATCAACATTGATACGA
Tbg972.6.4970      ------------------------------------------------------
```

Figure 5-12. DNA sequence alignment between Tb927.6.5180 and Tbg972.6.4970. The start and the end of the truncated *ingi* and full RIME insertions are coloured with red and highlighted in yellow respectively. The duplicated 12bp repeats are underlined with dashed lines and the potential PCR primers are in white colours in black background.

One set of the 12bp duplicated repeats is present in Tbg972.6.4790, which could be considered as molecular signature for MGE excision, and this could be useful in studying the phylogenetic inference of different *T.b.gambiense* isolates. The hypothesis is that when the PCR amplification is carried out using the proposed primers on *T.b.*

gambiense, the resultant amplicons would be easily recognisable bands resolved by agarose gel electrophoresis as the *T.b.brucei* amplicon will have a larger molecular weight due to the insertion of the RIME.

5.4 Discussion

To develop improved molecular epidemiological tools for the identification of *T. brucei* strains and the study of the epidemiology of sleeping sickness, a detailed, deeper, bioinformatics study into trypanosomal MGEs was required. Trypanosomatid sequencing projects have aided the investigation of MGEs and their functions. The dynamic nature of MGEs can cause a swift evolutionary divergence and this constitutes a potentially powerful tool for species identification and phylogenetic inference. Furthermore, MGEs can be used as passive markers for studying the processes of mutation and selection (Wickstead *et al.,* 2003).

Since the *RHS* genes and their pseudogenes contain hot spots for retroelement insertions, they were targeted in this study as being a suitable location in the genome to search for the insertions of elements, such as RIME and *ingi*, in the Tb927/4 strain genome. Initially 59 *RHS* genes were found in the database of which only 11 have ingi/RIME insertions. These are defined as pseudogenes. One of these *RHS* pseudogenes (Tb927.6.5180) has a double insertion. This study's findings were in contrast to those of Bringaud *et al.* (2002) where the authors reported on 55 *RHS* genes with 16 of them having an *ingi*/RIME insertion. Bringaud *et al.* (2002) reported only 1 full member of *RHS*6, and this did correlate accordingly with the findings in this study.

The increase in the number of identified *RHS*/pseudogenes could be due to the number of unclassified *RHS* genes founded in chromosome 7, or it may be due to gene duplication over the evolutionary process. As Emes and Yang (2008) have reported, the genes identified in *T.brucei* are frequently duplicated, suggesting interaction with the external host environment. In addition, the lower number of MGEs insertions and the truncated *ingi's* insertions could possibly be due to a higher rate of homologous recombination between repetitive DNA as Melville *et al.* (1999) have argued. Moreover,

according to the phylogenetic tree constructed from the amino acid sequence alignments of the *RHS* genes (Figure 4), there are three clades. These are *RHS*1 / *RHS*2, *RHS*3 / *RHS*4 and *RHS*5 / *RHS*6. These correlate with the families described by Bringaud *et al.* (2002); the authors have reported that *RHS*3 is related to *RHS*4 by 43% whereas *RHS*5 is similar to *RHS*6 by 50%.

Due to the abundance of *RHS*3 and *RHS* 4 subfamilies in TREU927 genome, they were chosen for detailed analysis in this study. This aims of this study were to investigate the mutational changes which have occurred in the *RHS*3 and *RHS*4 subfamilies, to categorise the *RHS* genes according to their evolutionary state, and finally to find a suitable marker to distinguish between *T.b.brucei* and *T.b.gambiense,* using the retro-analysis method. Initially two examples of mutational changes were investigated: firstly gene duplication, and the outcome of this was that the duplicated genes were chimeric genes, as a result of homologous recombination between two copies from different subfamilies, the crossing over taking place in the N-terminal region upstream of the retroelement insertion site (Bringaud *et al.,* 2002). These types of homologous recombination events are probably the tip of the iceberg, since several undetectable events probably occur between the abundant homologous sequences clustered in large sections of multiple chromosomes (Bringaud *et al.,* 2002). The founding of the fate selected gene duplications was sub-functionalization, since the duplicated genes exhibited the C-terminus of *RHS*3 including the Retrotransposon insertion site from the parent gene. Hurle (2004) has suggested that duplicated genes act as a source of protein sub-functionalisation, via evolution of key positions in the protein, and that this adaptation is reflective of organism-specific selective pressures. Also, the natural

selection estimation was calculated as ω=1, suggesting that early in their history, many gene duplicates experience a phase of comfortable selection or even accelerated evolution (Emes and Yang, 2008), which could be related to the parasite survival. Gene duplication has been shown to be critical to the survival of *P. falciparum* in the human host by aiding the development of drug resistance (Price *et al.,* 2004), and the antigenic diversity necessary for immune evasion (Freitas-Junior *et al.,* 2000). As Bringaud *et al.* (2002) have reported the diversity observed in the *RHS* multigene family is probably generated by the high rate of recombination in subtelomeric regions, and this may be advantageous for the parasite. A mutational example was studied in which the effect of the MGEs transposition in *RHS* 3 and *RHS*4 evolution was examined. Tb927.6.5170, which has a truncated *ingi* insertion, was studied, and chromosomal translocation was found to have occurred indirectly via homologous recombination intra-chromosomally.

The evolutionary time dimension of *RHS*3 and *RHS*4 in TREU927 genome was studied based on the presence of the duplicated 12bp repeat which has formed due to MGEs insertion. Since the original state and intermediate state were known, the recent state was investigated. We found the recent state was present in chromosome 1 (Tb927.1.70). The importance of this study can be interpreted as being to have fitted the three classes of the *RHS* evolutionary time dimensions into sleeping sickness epidemiological studies, since pseudogenes possess good properties as markers for epidemiological studies as they are large non-protein coding regions. As such they are subject to mutations that are free of selective pressure and are well suited for differentiating closely related isolates (Sacramento *et al.,* 1992). Based on the above classification of the time dimensions of evolutionary events of the *RHS* pseudogenes, a framework for differentiating between *T.brucei* isolates collected from different foci and

geographical locations could be developed. The principle behind this characterisation is that the state of the evolutionary timing of isolates (original, intermediate or recent) can be determined using this set of markers. So, for example, in a given collection of isolates, those that have evolved more recently can be distinguished from those that have evolved earlier. As Edwards and Gibbs (1992) have reported, even when a rare deletion of an element occurs, it leaves behind a molecular signature in the flanking sequences of a locus. This is could be useful for trypanosomiasis epidemiological studies, as it could allow us to differentiate between isolates from different foci and geographical locations according to temporal dimensions. Edwards and Gibbs (1992) have successfully used this approach whereby they reported on a comparison between a tightly linked marker within the human genome and studies of baboon DNA samples. This revealed that the original insertion was a relatively early event in primate evolution, but the *Alu* sequence loss leading to the dimorphism occurred much more recently. Loss of *Alu* insertions are therefore a mechanism for the generation of human *Alu* dimorphisms.

Finally, *ingi/*RIME (*T.brucei*), L1Tc/NARTc (*T.Cruzi*) and *L. major D*IREs share the first 78–79 nucleotides even though they are unrelated to each other (Bringaud *et al.,* 2006). It is thus hard to differentiate between the trypanosomatids using retroposons as markers. However, according to Jackson *et al.* (2010) intra-specific genome divergence between *T.b.brucei* and *T.b.gambiense* is estimated to be 1% where the indels or insertions are generally caused by mobile genetic elements. These authors have reported the deletion or insertion differences between *T.b.brucei* and *T.b.gambiense* being typically caused by transposable elements and internal VSG ‘islands’. On chromosome 9, a 7 kb insertion occurs in *T. b. brucei* due to an *ingi* element (at 1.24 Mb) not present in *T. b. gambiense.*

In this study we hypothesised that a RIME element could be inserted into one of the *RHS* genes of the TREU 927 genome but not in the correspondent *RHS* pseudogene in the DAL 972 (*T.b.gambiense*). We could thereby consider this as potential marker for differentiating between *T. brucei* and *T.b.gambiense*. Each *RHS*-psuedo gene that contains RIME insertion was compared to the DAL972 genome (BLAST search). It was found that *RHS* pseudogenes (Tb927.6.1150) have a high similarity (95%) with Tbg972.6.4970 *RHS* genes in term of the flanking region, but the RIME insertion was missing in the Tbg972.6.4970. Also, the fact that Tb927.6.1150 did not have similarities to the other *T.brucei* sub-species was an encouraging sign for the production of a reliable marker. However, due to the time limitations we were not able to test this provisional hypothesis, but past experiments including those of Stoneking *et al.* (1997), Witherspoon *et al*. (2006) and Monden*et al.* (2014) have indicated support for this hypothesis; the advantages of using MGEs as genetic markers is that they can show more accurate relationships than many other genetic markers such as single nucleotide polymorphisms (SNP), microsatellites, and restriction fragment length polymorphisms (RFLP).

MGEs including RIME play a very important role in phylogenetic studies of parasites, as well as in the epidemiology of sleeping sickness. Based on the findings of this chapter, a marker that could be used to differentiate between *T.b.brucei* and *T.b.gambiense* can be potentially developed.

6.1 General discussion

In this thesis ways to improve the field molecular diagnosis of African trypanosomiasis were studied. This study has centred around 36 blood samples of human patients, which were positively tested for trypanosomiasis by CATT, which were collected on FTA cards from Angola. Cox *et al*. (2005) have developed a nested ITS-PCR, which is more sensitive than the single round PCR reaction. This development allowed major simplification of the entire sample processing protocol whilst at the same time allowing the same diagnostic sensitivity to be applied in the detection of all important African trypanosome species. However, targeting a single punch from the FTA card matrices increase the chances of false negative PCR results since the targeted punch may lack parasite DNA (uneven distribution of the parasite DNA). A previous study that was carried out by Cox *et al*. (2010) has investigated the use of FTA cards in the diagnosis of trypanosomiasis. The authors have reported that a single punch from the FTA card may lead to false PCR negative results, and hence the wrong prevalence. This was found to be due to the uneven distribution of a sample on the card. Subsequently, this will have an impact on the sensitivity of this PCR technique. This is especially true when the parasite density is low, which means the chances of a single punch containing trypanosome DNA is low.

Therefore, initially in this study we have experimentally tackled this problem for diagnosis from the Angolan human samples. To achieve this, more than one punch (5 punches) from the patient samples that were collected from Angola, were selected for Chelex DNA elution prior to the ITS-PCR amplification. However, prior to the Chelex DNA elution, the validity of the Chelex elution was investigated using mammalian tubulin PCR. According to the mammalian tubulin PCR results the Chelex elution approach

could be considered as convenient and a reliable approach to be used in field diagnostic laboratories for trypanosomiasis. Also, the choice of the DNA extraction kit should be based on economic and time factors particularly in field laboratories located in Africa; considering the estimated cost of less than 0.16 US$ per sample for Chelex elution (Strøm *et al*., 2014). It can be concluded from this study Chelex elution has been shown to be reliable and economically suitable for trypanosomiasis molecular diagnosis. The advantages of Chelex elution includes low cost, less time and effort consuming, and other several advantages are mentioned in (chapter 3)

Moreover, the nested ITS-PCR technique efficiency in diagnosing HAT was tested. To begin with the hypothesis of the testing a single punch directly by nested ITS-PCR was studied. when single punch was taken from FTA spotted with trypanosome DNA, which have produced a 0% detection. However, as described in Chapter 3 (Figure 3-5) the advantage of the nested ITS-PCR over a single round ITS-PCR in terms of sensitivity was noticed. As the positive control (liquid trypanosome DNA) of the PCR reaction failed to produce any bands in the first reaction, but after the second reaction a clear band was recorded. That was produced by introducing to the two inner primers (ITS3 and ITS4), which makes the technique very much more sensitive, and in this application a two round amplification strategy has a second desirable effect. In the second round of the PCR the amount of the target is increased, whilst the amount of potentially inhibitory factors are greatly diluted (Cox, 2007). However, the nested Chelex/ITS-PCR has shown lower sensitivity when it was carried out on DNA eluted from FTA cards spotted with blood (Angolan Samples) than DNA eluted from FTA card spotted with liquid trypanosome DNA (*T. b. rhodesiense* stock from Uganda); 46.5% and 67% respectively. This could be attributed to different factors whether limitations related to the FTA card or to the assay itself. Technical problems that are related to the FTA card include; the age of the FTA cards, since the samples

were collected in the 2001 it is hard to remove all the PCR inhibitors, as Becker *et al.* (2004) have reported in assays, that long-term storage (>9 months) has been shown to prevent the complete removal of blood from discs by washing. Also, Matovu *et al*. (2010) attributed their low sensitivity of SRA-PCR, to a decrease in DNA quality during storage, as up to 1 year elapsed between sampling and execution of the tests.

In addition, the presence of the host DNA and blood inhibitors such as Haem, Lactoferrin, IgG and non-target DNA post the sample processing could contribute to the low sensitivity of the nested ITS-PCR in this particular study. As there are a variety of studies have proven that porphyrin compounds that are present in the haem component of the red blood cells are one of the major PCR inhibitory components of blood. These compounds have been shown to inhibit the activity of the thermo-stable DNA polymerases used in the PCR reaction (Loparev *et al*., 1991; Wooden *et al*., 1993; Akane *et al*., 1994; Cox, 2007). Although, the use of Chelex in DNA elution from blood samples helps in removing such inhibitors, by binding to the Chelex®100 bead matrix itself (Walsh *et al*., 1991). However, Becker *et al.* (2004) have found that the presence of PCR inhibitors cannot be completely ruled out after Chelex®100 extraction. A further evidence that PCR inhibitors could play role in lowering the sensitivity of the PCR amplification, Adams *et al.* (2006, 2008) reported a lower trypanosome detection using PCR (43.1%). The authors explained this low detection rate because the presence of inhibitors such as proteases or haem moieties from the midgut of the fly. However, the sensitivity was increased to 78% after storage of the midgut samples in 100% ethanol and purification of DNA using ammonium acetate precipitation.

Moreover, further technical issues in using FTA cards that may be related to the low sensitivity. Ahmed *et al*. (2013) have suggested that the low sensitivity of nested ITS-PCR may could be due to that the fact that the ITS reaction was optimized to work on washed discs rather than eluted DNA. As an explanation for the low sensitivity for Cox *et al*. (2005) ITS-PCR in comparison to TBR-PCR. However, Nijru *et al*. (2005) have reported the opposite, as the authors have extracted DNA from blood samples with Qiagen extraction kits that costs £3.02 per sample [Qiagen website, 2015] and 2 µl of the extract was used as a template in the PCR reaction. Extraction of DNA directly from blood samples increased the sensitivity for the ITS reaction while elution of DNA from the FTA cards inhibits the ITS PCR (Ahmed *et al.,* 2013).

On the other hand, the low sensitivity of the nested ITS-PCR could be attributed to the fact that internal transcribed genes present in low numbers ~hundreds, in comparison to 177 bp DNA satellite repeat sequence that present in large number ~10,000 in the *T.brucei* genome. As Thumbi *et al*. (2008) and Ahmed *et al*. (2013) explained their low sensitivity of nested ITS-PCR in comparison to TBR-PCR.

In summary, FTA card field application have many advantages which ease the collection, mobilization, and storage of the samples, and in this study have shown promising potential should further evaluation be considered. For instance by either lysing the blood before applying it on the FTA card that (Ahmed, 2009), MDA (Chapter 3) which costs nearly £13 per sample, or multiple washing before Chelex elution to ensure the removal of as much as possible of PCR inhibitors. Moreover, although nested ITS-PCR has offered great deal of help to the field diagnosis of trypanosomiasis in the past by reducing the consumed time and costs, and have shown a promising potentials in this study. Further evaluation of the technique is recommended. Particularly detailed investigation into the internal

transcribed genes in the pathogen *T.brucei* i.e. *T.b.gambiense* genome, in order to design more specific primers, which will enhance the sensitivity of the assay. Moreover, many publications including; Matovu *et al*. (2010) where the authors compared molecular trypanosome detection methods (nested SRA-PCR and SRA-LAMP [chapter1]). The authors have reported a higher sensitivity with less time consumed using SRA-LAMP than nested SRA-PCR, from a single punch from FTA card spotted with blood infected with tryapanosomiasis. Also, Ngotho *et al*. (2015) have tested the blood, cerebrospinal fluid (CSF), saliva, and urine, of six vervet monkeys were experimentally infected with *T. b. gambiense.* Using parasitological methods, *TgsGP* PCR and LAMP. The detection by LAMP was significantly higher than that of parasitological methods and PCR in all the samples. Therefore, in interest of comparing the sensitivity of ITS-PCR versus LAMP. Chelex/ITS LAMP should be taking into consideration, first by using the ITS primers from Cox *et al*. (2005), then with more specific primers should new primers been designed. Also, it would be interesting to find the outcome of Chelex/nested ITS-PCR, should it be carried on CSF samples rather than blood. To investigate its usefulness in staging the disease, and compare the its sensitivity between blood and CSF. Finally, nested Chelex/ITS-PCR and FTA cards promises the possibility of carrying out large-epidemiological studies on African trypanosomiasis in a simple and cost-effective way, which would contribute to the elimination of sleeping sickness road map.

Further objectives of this thesis were to investigate the sensitivity of the MGE-PCR compared to nested ITS-PCR on FTA cards and to evaluate the use of MGE-PCR directly from FTA cards. Previous studies have proven MGE-PCR is a powerful molecular tool in studying diseases caused by pathogenic parasites, such as

Toxoplasma gondii (Terry *et al*., 2001) and *T.brucei* (Tilley *et al*., 2003). According to our knowledge MGE-PCR was not tested on FTA cards. In this study MGE-PCR was carried out on eluted DNA from FTA cards of patients who have been tested positively for trypanosomiasis by CATT (Angolan samples). In this study MGE-PCR has recorded higher sensitivity than nested ITS-PCR in detecting trypanosomes. However, the specificity of REV B could not be determined through DNA sequencing, this could be due to nonspecific primer binding since the single primer protocol is used to amplify unknown region of DNA or REV B is not specific to detect trypanosome as it was proven in the successful amplification of tryp-negative blood. Also, the resultant data of the DNA sequencing was not readable i.e. mixed peaks and short nucleotides sequence. This could be due to the fact single primer amplification is a linear reaction, where the amount of the produced DNA post amplification as single primer will produce half the number of copies produced by two set of primers. This can be overcome by cloning as Matsui *et al* (1993) have reported; where the DNA will be cloned with either with either oligo (dT) priming or random hexamers, in the case of *T.brucei* mobile elements where the 3' end is occupied by poly (A) tail oligo (dT) is more suitable. The produced population of cDNA will be then amplified by single primer amplification. Subsequent the amplification the amplified products will be inserted into suitable bacteriophage (DNA cloning).

In order to improve the sequencing of the amplified products by REV B, the produced amplicons were cut by restriction enzyme (AlwNI) to ease the priming process for the primer. However, the DNA sequencing of the post-amplification product did not record any sign of improvement. Which may due to the presence of contaminants that cannot be separated from the template DNA that led to inefficient amplification by REV B.

Which could also have affected the sequencing of RIME A and RIME B amplified products.

Although MGE-PCR recorded a higher sensitivity than nested ITS-PCR. However, the sequencing data of the amplicons resulting from REVB was not readable. Even though different approaches were followed (chapter 4) to improve the sequencing data, yet with no sign of improvement. Thus, in terms of future work another approach could be followed to improve the sequencing data, subsequently determining the specificity of REV B. Which is generating double-stranded DNA from the linear reaction caused by single primer amplification. By oligo (dT) priming, once the double stranded DNA is generated, it will be cloned into a suitable bacteriophage. After growing sufficient amount DNA, sequencing the product may give a clearer picture about the specificity of the REV B as Matsui *et al*. (1993) have reported.

Based on the findings in (chapter 4), a more specific protocol based on MGEs is required. To achieve this, a bioinformatics analysis of mobile elements inserted in RHS/pseudogenes in the *T.brucei* genome was carried out. The aims of this were to locate variable regions of these genes that could be used as detailed markers for trypanosome strain identification to improve the specificity of MGE-PCR. Sequences of the RHS genes were retrieved from the *T.brucei brucei* and *T. brucei gambiense* genomes to investigate positional diversity of MGEs within this family of genes. Differences were found in the presence/absence of RIME elements in one RHS gene between the two subspecies. More detailed investigation of all RHS gene classes in *T.b.brucei* showed six classes of RHS gene types and within each class, individual sequences showed evidence of insertion by MGEs.

In some specific instances, evidence of pre-insertion, insertion and subsequent removal of MGEs was seen. This enabled a temporal evolutionary sequence of events to be interpreted.

We were able to class the RHS pseudogenes into three categories including; the original state, intermediate state and the recent state. Based on the presence and the excision of the mobile element. This will aid the study of sleeping sickness epidemiology, by tracking the migration of *T.brucei* populations, according to if the elements are present in some species and not others, they can be used to determine the evolutionary history of those species. As Batzer and Deininger, 2002 reported an *Alu* subfamily known as young *Alu* subfamilies (have inserted into the human genome most recently) are polymorphic with respect to the presence or absence of insertion in different human genomes. Those relatively few elements that are present in the genomes of some individuals and absent from others are referred to as Alu-insertion polymorphisms. This polymorphism has been used to study human genomic diversity (Roy-Engel *et al*., 2001).

A potential marker that would differentiate between the *T.b.brucei* and *T.b.gambiense* was found. The potential of this marker is high as it would differentiate between the two sub-species via the produced amplicons. Where an insertion of RIME was found in the TREU927 genome and was absent in the Tbg972 genome, this represents a potential mobile element marker for studies of the epidemiology and control of sleeping sickness. Where two PCR primers from the flanking region of the RIME (chapter 5), rather than single primer (REV B). That will amplify the RIME present in the parasite genome, in case of *T.b.gambiense* the produced gel band will be smaller than *T.b.brucei*, due to the absence of the RIME insertion in the *T.b.gambiense*. Hypothetically based on the bioinformatics experiments, this is considered as powerful tool that will ease studying large-scale samples, and

economically friendly. As the differentiation could be visualised by the amplicon sizes on the gel image. Most studies of population genetics in *T.brucei* species are facilitated by mitochondrial DNA, microsatellites, AFLP, or RFLP. However, retrotransposon insertions represent a valuable tool because of their unique combination of genetic properties and the observation that they are one of the least expensive molecular markers to assay. Ideally, all one needs to assay a population is a thermal cycler and gel electrophoresis equipment. This approach will allow for different *T.brucei* isolates to be typed in large- scale studies. These genes offer the opportunity to develop specific molecular epidemiological tools for investigating the evolution of MGEs in field samples and the possibility of a method for temporally informed epidemiological tracking of isolates.

One of the limitations of this study is the number of the studied RHS genes families (RHS3 and RHS4 only), in order to make this study more specific and accurate is to study the rest of the RHS gene families' i.e. RHS1, RHS2, RHS5, and RHS6. Also, in this study only three genomes of the *T.brucei* subspecies have been studied which are TREU 927, Tb427 and Tbg972, which belong to *T.brucei, T.brucei* (lister strain) and *T.b.gambiense* subspecies respectively. To take this approach forward in the future it would be useful to conduct a similar investigation to the 1092 human genome project where scientists have planned to sequence the genomes of at least one thousand anonymous human participants from a number of different ethnic groups within three years (McVean *et al*., 2012). Such a project in trypanosomes could open a large number of possibilities for the development of new tools for the detection and tracking of trypanosomes. It, of course, would not be achievable before the 2020 proposed deadline for elimination.

In summary, these studies should also shed new light on the biology of these interesting mobile elements and provide a comparative assessment of their role in shaping the genetic relationship between *T.brucei* species.

Moreover, the meaning of elimination is “reduction to zero of the incidence of infection caused by a specific pathogen in a defined geographical area, as a result of deliberate efforts; continued actions to prevent re-establishment may be required.” (WHO, 2013). According to the past, HAT elimination was proven to be a hard task and requires huge amount of effort, time, and resources (Franco *et al*., 2014b). Therefore, there are questions to be asked as we are five years away from the deadline that was set by WHO in 2013. First are we in position to meet this target? Second should we meet the deadline, is it going to be sustainable elimination? WHO (2013) have designed this road map into 2 phases; the initial goal is 2020 where the predicted reported cases to be one new case per 10 000 inhabitants in 90% of foci. The final goal in the future should be the sustainable elimination of the disease, defined as the interruption of the transmission of gambiense HAT, which has been targeted for 2030 (WHO, 2013).

WHO (2013) are optimistic and believe this road map is feasible, should several measures be followed; one of the steps is developing new tools for case detection, that will extend the possibilities for diagnosis and surveillance in the health system. This present study concentrates on improving the current molecular diagnostic tools and developing a new marker that has the potential to contribute towards meeting the target by 2020. Chelex/nested ITS-PCR is considered a reliable tool for detecting trypanosomes in the field. However, the outcome of this study suggests that various improvements are required. For instance the usefulness of FTA cards as a source for storing and mobilizing the samples. Previous studies and this study have shown that the age of the stored sample on FTA card has a negative impact on the detection

rate by molecular tools, in this case the nested ITS-PCR. Also, the form of the samples, as in this study whole blood was shown that have a negative impact on the detection rate by nested ITS-PCR. Due to the PCR inhibitory factors present in the blood. Even though for the samples who have been positively tested for DNA elution using mammalian tubulin PCR, have failed to amplify after the second round nested ITS-PCR. This raise a questions related to the validity of FTA cards in field diagnosis of trypanosomiasis. Further concern that should be taking into consideration is to test CSF rather than blood, as especially in the Gambian form of HAT, the parasite density will be present in higher concentration in CSF than in the blood, particularly at the later stages of the Gambian HAT. Kyambadde *et al.* (2008) have reported a higher sensitivity when the PCR was used to detect trypanosome in the CSF (61.7%) than in the blood (57.1%). Also, in this study patients were recorded as second-stage of the Gambian HAT, were missed by the nested ITS-PCR amplification. This highlights the shortage of the assay in determining the stage of the disease. The outcome of this study have agreed with the weakness of the molecular diagnostic tools (nested ITS-PCR) against the parasitological tools (CATT) in diagnosing Gambian sleeping sickness. Also, Njiru *et al*. (2004) have reported a comparison between three methods including CATT, PCR and MHCT, in detecting *T.evansi* in 549 camels from Kenya. The authors have reported that CATT is more sensitive in determining the prevalence of trypanosomes (45.5%) using CATT, in comparison to (5.3%) and (26.6%) for MHCT and PCR respectively. However, in previous studies (Penchenier *et al*., 2000; Solano *et al*. 2002) have reported that molecular tools are more sensitive using TBR PCR Moser *et al*. (1989), than CATT.

On the other hand, Solano *et al*. (2002) have reported 100% sensitivity and 92% specificity, when Chelex was used to elute DNA from blood samples, prior TBR PCR. This support the suggestions of Ahmed *et al*. (2013), that nested ITS-PCR developed by Cox *et al*. (2005) was only designed to work on a single punch from the FTA cards. Beside, according to the outcome of (chapter 4) where Chelex/MGE-PCR have shown higher sensitivity than Chelex/ITS-PCR from the same samples. Yet the MGE-PCR lacks the specificity based on the post-amplification analysis, to investigate the produced amplicons was unsuccessful. However, it was the first time (to our knowledge) that MGE-PCR was applied directly on FTA cards spotted with blood of infected patients by trypanosomiasis. Yet it has shown promising potentials, should more efforts be carried to increase the specificity of the MGE-PCR. Which have led us to investigate the mobile elements in *T.brucei* genome in more depth bioinformatically. In order to develop a new molecular marker based on the MGEs present in *T.brucei* genome and strengthen the prospect of MGE-PCR.

The elimination of gambiense HAT is largely relying on the sustained political stability in Africa. Welburn and Maudlin (2012) have expressed their views on the role of the public-private funding in eliminating gambian sleeping sickness. The authors have suggested that public–private funding mechanisms should not be excluded. It is timely to take up the tools available and invest in new tools – including novel financial instruments – to eliminate this disease from Africa. Also, despite the epidemiology of gambiense HAT being generally well understood, some scientific questions remain unanswered. Particularly in areas related to the presence and distribution of the disease, as well as the role of animals as reservoirs of *T. b. gambiense* (WHO, 2013).

As Njiokou *et al*., (2010) have reported that a PCR results have detected *T. b. gambiense* group 1 DNA in 27 (3.08 %) domestic animals of 875 animals from Cameron. If *T. b. gambiense* is able to be transmitted from animals to humans and from humans to animals, it will have an impact on the transmission rate, as it will increase the transmissibility index. Subsequently the expected number of reported cases will be increased as well.

The outcome of this study (chapter 5) a potential marker to be used in strain typing of *T.brucei* has been developed. We believe that trypanosomiasis epidemiological studies will undoubtedly benefit from this development. The principle of this marker is that isolates of *T.brucei* from different geographical locations and periods can be differentiated, according to the insertion or the absence of the MGEs (RIME) in the genome. This molecular marker will answer epidemiological related question, particularly in cases that are falling outside tsetse-infested areas, cases originating from areas with no reported history of HAT. Moreover, since HAT is described as focal disease. Therefore, Geo-referenced data are crucial in controlling programmes of HAT. As we learnt from the "Dracunculiasis Eradication Program", where the number of cases has dropped from 3.5 million in the mid-1980s in 20 countries worldwide, to 126 in 2014 (Weekly epidemiological record, 2010; Global Health Observatory, [WHO website]), where the Geo-referenced has played an important role. This potential marker can benefit from Geographical Information Systems (GIS) -based methodology for risk estimation was developed and tested in six Central African countries (Cecchi *et al*., 2009). Since an ideal GIS would contains detailed information on each known case of the disease, including: geographic coordinates of the patient's household, sex and age of the patient, exact date of disease detection or reporting, duration of illness, mortality/recovery rates. Also, information on the location of the patient's working place would be needed, as well as the patterns of people mobility in

the area (Cecchi *et al*., 2009). Such information will be very useful for predicting the evolution of the disease and the efficiency of the control measures. Also, GIS will enable us to identify, localize and characterize biotopes and zones of high transmission risk where vector control can be implemented. Consequently, vector control costs will be reduced (Simo *et al*., 2014). Thus, a combination of work between the mapping methods and the new potentially developed marker would have a huge positive impact towards the road map of HAT elimination.

In summary, the road map to eliminate HAT by 2020, and the decline in the number of cases, reflects the increase awareness of the disease and the huge amount of efforts and work done over the years, in developing new tools to control sleeping sickness. In this study we present ways to improve molecular diagnosis of sleeping sickness, and it can be concluded that further investigations are required to compete with the complexity of the disease and the pathogens. Also, this study produced a new potentially promising genetic marker to be used in epidemiological studies of trypanosomiasis in the future.

6.1 Future work

The outcome of this research is promising and worthy of further development. Firstly, an in-depth investigation into the internal transcribed spacer genes within the *T.b.gambiense* genome is warranted in order to design enhanced species specific primers that could be tested on the Angolan patient samples. Secondly, the Chelex eluted DNA from the Angolan patient FTA cards could be tested using the TBR-PCR approach. This should give a clearer picture regarding the presence, or the absence, of trypanosome DNA since TBR-PCR has been shown to have a higher sensitivity than ITS-PCR in detecting trypanosomes. Finally, an alternative molecular tool, LAMP (see Chapter 1) could be applied on the eluted DNA from the FTA cards that were collected from Angola, in order to compare the sensitivity between ITS-PCR and LAMP techniques. Furthermore, it would be useful to collect additional blood samples from healthy participants from Angola. These could be analysed by the nested ITS-PCR approach in order to have greater confidence in the specificity of the technique when applied to the detection of trypanosomes. Finally, based on the bioinformatics results presented in Chapter 5, the designed primers should be experimentally applied on the Angolan samples

6.2 Conclusion

Successful molecular diagnosis of sleeping sickness in Africa is largely based on the availability of the field laboratories that require a genomic material of an appropriate quality and concentration to be present within the sample under examination. Thus, according to the work presented in this thesis, different blood preparations should be considered before spotting onto FTA cards. Also, a further consideration should be taking into account, is the age of the FTA card. From this study based on economical and time consuming considerations Chelex elution is highly recommended to extract genomic materials from the FTA card. Moreover, Chelex/ITS-PCR is arguably economically acceptable and less time consuming, however, further investigations are required to improve its sensitivity. Furthermore, despite the successful history of MGE-PCR in epidemiological studies of sleeping sickness, based on conclusion of chapter 4 REV B specificity is still under question. Finally, from the bioinformatical experiments chapter 5, where the use of mobile elements in the diagnosis of trypanosomiasis was investigated. It can be concluded from the retro-analysis study, mobile elements as markers can possess a great potential for studying the epidemiology of sleeping sickness, which will enhance the prospects of MGE-PCR in term of specificity.

7. References

Abel. P, Kiala. G, Lôa. V, Behrend. M, Musolf. J, Fleischmann. H, Théophile. J, Krishna. S and Stich .A (2004). Retaking sleeping sickness control in Angola. *Tropical Medicine & International Health*, **9**(1). 141–148.

Adams. E and Hamilton. B (2008). New molecular tools for the identification of trypanosome species. *Future Microbiology*. **3**(2):167-76.

Adams. ER, Malele. II, Msangi. AR, Gibson. WC (2006) Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. *Act Trop*. **100**:103-109.

Agbo. E, Majiwa. P, Claassen. H, te Pas. M (2001). Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. *Parasitology.* **124**: 349-358.

Ahmed. H (2009). African trypanosomes; a model for the improvement of molecular diagnosis of blood borne parasites. PhD Thesis. University of Edinburgh

Ahmed. H, MacLeod. E, Hide. G, Welburn. S and Picozzi. K (2011) The best practice for preparation of samples from FTA® cards for diagnosis of blood borne infections using African trypanosomes as a model system. *Parasites & Vectors*. **4**:1-7.

Ahmed. H, Picozzi. K, Welburn. S, Macleod. E (2013). A comparative evaluation of PCR- based methods for species- specific determination of African animal trypanosomes in Ugandan cattle. *Parasites and Vectors*. **6**:316.

Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *Journal of Forensic Science*. **39**; 362-372.

Aksoy. S (2003). Control Tsetse Flies and Trypanosomes Using Molecular Genetics.

Veterinary Parasitology, **115**: 125-145.

Aksoy. S, Williams. S, Chang. S and Richards. F. F. (1990). SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Research* 18(4); 785-792.

Al-Soud, W.A (2000). Optimisation of diagnostic PCR a study of PCR inhibitors in blood and sample pre-treatment. Lund University, Sweden.

Altman. D (1991). Practical statistics for medical research. London: Chapman and Hall.

Archie. E., Luikart. G., Ezenwa. V (2008). Infecting epidemiology with genetics: a new frontier in disease ecology. *Cell* **24** (1). 21-30.

Attardo. G. M *et al* (2014**).** Genome Sequence of the Tsetse Fly (Glossina morsitans): Vector of African Trypanosomiasis**.** *Science*, **344** (6182): 380-386.

Avelar. D and Linardi. P (2010). Use of multiple Displacement Amplification as Pre-Polymerase Chain Reaction (PRE-PCR) to Amplify Genomic DNA of siphonapterids preserved for a Long periods in scientific collections. *Parasites and Vectors.* **3**; 86.

Balmer. O, Beadell. JS, Gibson. W and Caccone. A (2011) Phylogeography and taxonomy of *Trypanosoma brucei*. *PLoS Neglected Tropical Diseases*. **5**:e961.

Balmer. O, Palma. C, MacLeod. A and Caccone. A (2006). Characterization of di-, tri-, and tetranucleotide microsatellite markers with perfect repeats for *Trypanosoma brucei* and related species. *Molecular Ecology Notes*, ***6***(2), 508–510.

Barrett. M. P, Burchmore. R, Stich. A, Lazzari. J. Frasch. A, Cazzulo. J and Krishna. S (2003). The Trypanosomiases. *THE LANCET*, **362**: 1469-1480.

Barrett. M. P. (1999). The fall and rise of sleeping sickness. *THE LANCET*, **553**: 1113-114.

Batzer. M and Deininger. P (2002). *Alu* repeats and human Genomic diversity. *Nature reviews genetics*. **3**:370-379.

Becker. S, Franco. J, Simarro. P, Stich. A, Abel. P and Steverding. D (2004). Real-time PCR for detection of *Trypanosoma brucei* in human blood samples. *Diagnostic microbiology and infectious diseases*. **50** (3): 193-199.

Berberof. M, Pérez-Morga. D and Pays. E (2001). A receptor-like flagellar pocket glycoprotein specific to *Trypanosoma brucei gambiense. Molecular and Biochemical Parasitology.* **113**(1):127-138.

Bhattacharya. S, Bakre. A, Bhattacharya. A (2002). Mobile genetic elements in protozoan parasites. *Journal of Genetics*. **81**(2):73-86.

Biteau. N. Bringaud.F, Gibson.W, Truc. P and Baltz. T (2000). Characterization of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. *Molecular and Biochemical Parasitology*. **105**:185-201.

Blaxter. M (2001). Parasite genomes in; Mar. J, Nilsen. T and Komuniecki. R *Molecular Medical Parasitology* MPG Books, Bodmin, Cornwall; 3-27.

Blum J, Neumayr. A and Hatz C. (2011). Human African trypanosomiasis in endemic populations and travellers. *Europe Journal of Clinical Microbial Infectious Diseases*. **31**(6):905-13.

Brindley. P, Laha. T, McManus. D and Loukas. A (2003). Mobile genetic elements colonizing the genomes of metazoan parasites, *Trends in Parasitology*, **19** (2); 79-87.

Bringaud. F, Biteau. N, Berriman. M, El-Sayed. N, Ghedin. E, Melville. S. E, Hall. N, Baltz. T, Hez. S and Leech. V (2002). A New, Expressed Multigene Family Containing a Hot Spot for Insertion of Retroelements Is Associated with

Polymorphic Subtelomeric Regions of *Trypanosoma brucei Eukaryotic Cell.* **1** (1): 137-151.

Bringaud. F, Biteau. N, Zuiderwijk. E, Berriman. M, El-Sayed. M Ghedin. E, Melville. S. E, Hall. N and Baltz,T (2004). The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei. Molecular biology and evolution.***21** (3); 520 - 520

Bringaud. F, Ghedin. E, Najib. A, El-Sayed. N and Papadopoulou (2008) Role of transposable elements in trypanosomatids. *Microbes and Infection* **10;** 575-581

Bringaud. F, Ghedin. F, Blandin. G, Bartholomeu. D, Caler. E, Levin. M.J, Baltz. T and El-Sayed. N (2006). Evolution of non-LTR retrotransposons in the trypanosomatid genomes: Leishmania major has lost the active elements. *Molecular and Biochemical Parasitology*. **145**:158-170.

Bringaud. F, Müller. M, Cerqueira. G. C, Smith. M, Rochette. A, El-Sayed. N. M. A and Ghedin. E (2007). Members of a Large Retroposon Family Are Determinants of Post-Transcriptional Gene Expression in *Leishmania*. *PLoS Pathogens*, **3**(9), e136.

Bromidge. T, Gibson. W, Hudson. K and Dukes. P (1993). Identification of *Trypanosoma brucei gambiense* by PCR amplification of variant surface glycoprotein genes. *Acta Trop* **53**, 107-119.

Brun. R, Blum. J, Chappuis. F, and Burri. C (2010). Human African trypanosomiasis. *The Lancet*. **375** (9709):148-159

Bukachi. S, Wandibba. S, and Nyamongo. K (2009). The Treatment Pathways Followed by Cases of Human African Trypanosomiasis in Western Kenya and Eastern Uganda. *Annals of Tropical Medicine and Parasitology.***103** (3):211-20.

Bull. JJ (1994). Perspective: Virulence. *Evolution* **48**(5): 1423–1437.

Büsher. P and Lejon. V (2004). *Diagnosis of human African trypanosomiasis.* In*:* The Trypanosomiasis (Maudlin, I et al., eds) pp 203-218, CABI Publishing*.*

Cecchi. G, Paone. M, Franco. J. R, Fèvre. E. M, Diarra. A, Ruiz. J. A and Simarro. P (2009). Towards the Atlas of human African trypanosomiasis. *International Journal of Health Geographics.* **8:**15.

Chappuis. F, Lima. A, Flevaud. L and Ritmeijer. K (2010). Human African Trypanosomiasis in Areas without Surveillance. *Emerging Infectious Diseases*, **16**:354–356.

Chappuis. F, Loutan. L, Simarro. P, Lejon. V and Büscher.P (2005). Options for Field Diagnosis of Human African Trypanosomiasis. *Clinical microbiology reviews* **18** (1): 133-146.

Checchi. F, Filipe. J, Barrett. M, Chandramohan. D (2008a). The natural progression of gambiense sleeping sickness: what is the evidence? *PLoS Neglected Tropical Diseases*, **74**(12); e303.

Checchi. F, Filipe. J, Haydon. D, Chandramohan. D and Chappuis. F (2008b). Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infect Dis*.**8**:16.

Chen. R, Tong. Q, Zhang. Y, Lou. D, Kong. Q, Lv. S. Z. M, Wen. L and Lu. S (2011). Loop-mediated isothermal amplification: rapid detection of *Angiostrongylus cantonensis* infection in *Pomacea canaliculata*. *Parasites and Vectors.* **4** (204):1-7.

Clausen. P, Wiemann. A, Patzelt. R, Kakaire. D, Poetzsc. C, Peregrine. A and Mehlitz. D (1998). Use of a PCR assay for the specific and sensitive detection of

Trypanosoma spp. in naturally infected dairy cattle in peri-urban Kampala, Uganda. *Annals of the New York Academy of Sciences.* **849**: 21-31.

Control and surveillance of African trypanosomiasis (1998). Report of a WHO Expert Committee. Geneva, World Health Organization, (WHO Technical Report Series, No. 881).

Cordon-Obras. C, Berzosa. P, Ndong-Mabale. N, Bobuakasi. L, Buatiche. JN, *et al* (2009). *Trypanosoma brucei gambiense* in domestic livestock of Kogo and Mbini foci (Equatorial Guinea). *Tropical Medicine & International Health*. **14**:535–541.

Cox. A (2007). Epidemiological analysis of host populations with widespread of sub-patent infections. PhD thesis. Edinburgh University, Edinburgh.

Cox. A, Tilley. A, McOdimba. F, Fyfe. J, Hide. G and Welburn. S (2005)**.** A PCR based assay for detection and differentiation of African trypanosome species in blood. *Experimental Parasitology*, **111,** 24-29.

Cox. A, Tosas. O, Tilley. A, Picozzi. K, Coleman. P, Hide. G and Welburn. S (2010) Constraints to estimating the prevalence of trypanosome infections in East African zebu cattle. *Parasites & Vectors*. **3**: 1-8.

Crisione. C, Poulin. R and Blouin. M (2005). Molecular ecology of parasites: elucidating ecological and evolutionary processes. *Molecular Ecology.* **14**: 2247-57.

Czerny. T (1996). High primer concentration improves PCR amplification from random pools. *Nucleic Acids Research*. **24**(5): 985–986.

Daffa. J, Byamungu. M, Nsengwa. G, Mwambembe. E and Mleche. W (2013). Tsetse distribution in Tanzania: 2012 status. *Tanzania Veterinary Journal.* **28**: 1-11.

De Clare Bronsvoort. B, von Wissmann. B, Fèvre. E, Handel. G, Picozzi. K, and Welburn. S (2010). No gold standard estimation of the sensitivity and specificity of two molecular diagnostic protocols for *Trypanosoma brucei* spp. in Western Kenya. *PLoS One*. **5**:e8628.

De Raadt. P (1999). The history of sleeping sickness. In: Gilles HM, ed. protozoal diseases. London, Arnold.

De Vooght. L, Caljon. G**,** Stijlemans. B**,** Baetselier. P**,** Coosemans. M and Den Abbeele. J (2012). Expression and extracellular release of a functional anti-trypanosome Nanobody® in *Sodalis glossinidius*, a bacterial symbiont of the tsetse fly. *Microbial Cell Factories* **11**: 1-11.

Deborggraeve. S and Büscher. P (2010). Molecular diagnostics for sleeping sickness: what is the benefit for the patient?. *The Lancet Infectious Disease*. **10** (6) 433-439.

Despommier. D, Gwadz. R, Hotez. P and Knirch. C (2005). African Trypanosomiasis. *Parasitic Diseases*. New York: Apple Trees Productions LLC. Ch 6.

Desquesnes. M, Mclaughlin. G, Zoungrana. A and Dávila. A (2001). Detection and Identification of *Trypanosoma* of African Livestock through a Single PCR Based on Internal Transcribed Spacer 1 of rDNA. In*ternational Journal for Parasitology*. **31**: 610–614.

DND*i* (2014). R&D status September 2014: DNDi Human African trypanosomiasis programme. [Online] [Accessed on 12/1/15] available: http://www.dndi.org/media-centre/news/1960-hat_rndstatus_092014.html

Donelson. J (1996). Genome Research and Evolution in Trypanosomes. *Current Opinion in Genetics and Development*. **6;** 699-703.

Dorman.J.S (1992). Genetic Epidemiology of insulin-dependent diabetes mellitus: International Comparisons Using Molecular Genetics. *Annual Medicine*. **24**: 393-399.

Dukes. P, Gibson. W, Gashumba. J, Hudson. K, Bromidge. T, Kaukus. A, Asonganyi. T and Magnus. E (1992). Absence of the LiTat 1.3 (CATT antigen) gene in *Trypanosoma brucei gambiense* stocks from Cameroon. *Acta Tropica*. **51**:123-134.

Dyer. NA, Rose. C, Ejeh. NO and Acosta-Serrano. A (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends in Parasitology*. **29**(4):188–196.

Ebert. D and Herre. E (1996). The evolution of parasitic diseases. *Parasitology Today* **12**: 96–101.

Edwards. M, Gibbs. R (1992). A human dimorphism resulting from loss of an Alu, *Genomics*, **14**(3):590-597.

Eisler. M.C, Dinger. R.H, Majiwa. P.A.O and Picozzi, K (2004). Diagnosis and epidemiology of African Animal Trypanosomiasis, In: The Trypanosomiasis. Maudlin, I., Holmes, P. H. & Miles, M. A. Ed., Oxford, UK.

Ekwanzala M *et al*. (1996). In the heart of darkness: sleeping sickness in Zaire. *Lancet*, **348**:1427–1430.

El-Sayed. N *et al* (2005). The sequence and analysis of *Trypanosoma brucei* Chromosome II*. Nucleic Acid Research.* **31** (16); 4856-4863.

El-Sayed. N, Hegdea. P, Quackenbusha. J, Melvilleb. S and Donelsonc. J (2000). The African Trypanosome Genome. *International Journal for Parasitology*. **30**:329-345.

Elsen. P, Amoudi. MA and Leclercq M. A (1990). Discovery in Saudi Arabia of two species of tsetse fly that are vectors of human and animal trypanosomiasis. *Revue Médicale de Liège* [Article in French], **46**:225–231.

Eperon. G, Balasegaram. M, Potet. J, Mowbray. C, Valverde. O and Chappuis. F (2014). Treatment options for second-stage gambiense human African trypanosomiasis. *Expert Review of Anti-infective Therapy*. **12**(11): 1407–1417.

Fairlamb. A (2003). Chemotherapy of human African trypanosomiasis: current and future Prospects. *TRENDS in Parasitology*. **19** (11): 488-494.

Fevre. E, Coleman. G, Odiit. M, Magona. J, Welburn. C and Woolhouse. M (2001). The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *Lancet*, **358**: 603–604.

Fèvre EM, Wissmann Bv, Welburn SC, Lutumba P (2008). The Burden of Human African Trypanosomiasis*. PLoS Negl Trop Dis* **2(**12): e333.

Figueiredo. L, Pirrit. L and Scherf. A (2000) Genomic organisation and chromatin structure of *Plasmodium falciparum* chromosome ends. *Molecular and Biochemical Parasitology*. **106**: 169–174.

Forde. R (1920). Some clinical notes on a European patient in whose blood a trypanosome was observed. *Journal of Tropical Medicine* **5**: 261–263.

Franco. J, Simarro. P, Diarra. A and Jannin. J (2014a). Epidemiology of human African trypanosomiasis. *Clinical Epidemiology*. **6**: 257—275.

Franco. J, Simarro. P, Diarra. A, Ruiz-postigo. J and Jannin. J (2014b). The journey towards elimination of gambiense human African trypanosomiasis: not far, nor easy. *Parasitology*. **141**(6): 748-760.

Gautret. P, Clerinx. J, Caumes. E, Simon, F, Jensenius. M, Loutan. L, Schlagenhauf.

P, Castelli. F, Freedman. D, Miller. A, Bronner. U and Parola. P (2009). Imported Human African Trypanosomiasis In Europe, 2005-2009. *Eurosurvellince* **14** (36) 1-3.

Gibson, W (2002). Epidemiology and diagnosis of African trypanosomiasis using DNA probes. *The Royal Society of Tropical Medicine and Hygiene* **96** Suppl 1, S141-143.

Gibson. W (1986) Will the real *Trypanosoma b. gambiense* please stand up. *Parasitology Today*. **2**: 255–257.

Gibson. W (1989). Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei brucei*. *Parasitology* **99**: 391–402.

Gibson. W (2007). Resolution of the species problem in African trypanosomes. *International Journal of Parasitology*. **37**:829–838.

Gibson. W, Backhouse. T and Griffiths. A (2002). The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. *Infection, Genetics and Evolution.* **1:** 207–214.

Gibson. W, de C Marshall. T and Godfrey. D (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus Trypanozoo. *Advances in Parasitology*. **18**:175-246.

Gibson. W, Stevens. J and Truc. P (1999). Identification of trypanosomes: from morphology to molecular biology. In: "Progress in Human African Trypanosomiasis sleeping sickness", Editors M. Dumas, B. Bouteille, A. Buguet, Springer Publ., Paris, pp. 7-29.

Gobbi. F and Bisoffi. Z (2012). Human African trypanosomiasis in travellers to Kenya. *Euro surveillance.* **17**(10): 2-3.

Goodwin. LG (1985). Trypanosomiasis. Introduction. *British Medical Bulletin* **41**(2): 103-104.

Gray. Y (2000). Transposable-element-mediated chromosomal rearrangements. *Trends in Genetics*. **16**(10):461-468.

Hamilton. P, Adams. E, Malele. I and Gibson. W (2007). A novel high throughput technique for species identification, reveals a new species of tsetse-transmitted trypanosome related to *Trypanosoma brucei*. *Infection, Genetics and Evolution*. **8**(1):26-33.

Hartl. D, Lohe. A and Lozovskaya. S (1997). Modern thoughts on an ancient mariner: function, evolution, regulation. *Annu. Rev. Genet*. **31**: 337–358.

Hasan. G, Turner. M and Cordingley. J (1984). Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei. Cell.* **37**:333–341.

Hernandez. P, Martin-Parras. L, Martinez-Robles. L and Schvartzman. J (1993). Conserved features in the mode of replication of eukaryotic ribosomal RNA genes. *The EMBO Journal*. **12**(4): 1475–1485.

Hide, G and Tait, A (2009) Molecular epidemiology of African sleeping sickness. *Parasitology*.**136***:*1491 - 1500.

Hide. G, Angus. S, Holmes. P, Maudlin. I and Welburn. S (1998). Comparison of *Trypanosoma brucei* strains circulating in an endemic and an epidemic sleeping sickness focus. *Experimental Parasitology*. **89**: 21-29.

Hide. G, Buchanan. N, Welburn. S, Maudlin. I, Barry. I and Tait. A (1991). *Trypanosoma brucei rhodesiense*: characterisation of stocks from Zambia, Kenya and Uganda using repetitive DNA probes. *Experimental Parasitology*. **72**(4): 430–439.

Hide. G, Buchanan. N, Welburn. S, Maudlin. I, Barry. J and Tait. A (1991). *Trypanosoma brucei rhodesiense*: characterisation of stocks from Zambia, Kenya, and Uganda using repetitive DNA probes. *Experimental Parasitology*. **72**: 430–439.

Hide. G, Tilley. A, Welburn. S, Maudlin. I and Tait. A (2000). *Trypanosoma brucei*: identification of trypanosomes with genotypic similarity to human infective isolates in tsetse isolated from a region free of human sleeping sickness. *Experimental Parasitology*. **96**: 67 – 74.

Hide. G, Welburn. S, Tait. A and Maudlin. I (1994). Epidemiological relationships of *Trypanosoma brucei* stocks from south east Uganda: evidence for different population structures in human infective and non-human infective isolates. *Parasitology*. **109**:95-111.

Hide. G. (1999). History of Sleeping Sickness in East Africa. *Clinical Microbiology Review*, **12**(1): 112-125.

Hide. G. and Tilley. A (2001). Use of mobile genetic elements as tools for molecular epidemiology. *International Journal of Parasitology*. **31**: 599-602.

Hoare. C. (1972). *The trypanosomes of mammals. A zoological monograph.* Oxford: Blackwell. 750.

Hofreiter. M, Jaenicke. V, Serre. D, Haeseler. Av and Pääbo. S (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*. **29**:4793-4799.

Human African trypanosomiasis (sleeping sickness): epidemiological update (2006). *Weekly Epidemiological Record.* **81**:71–80.

Jackson. A *et al* (2010). The Genome Sequence of *Trypanosoma brucei gambiense*, Causative Agent of Chronic Human African Trypanosomiasis. *PLOS Neglected Tropical Diseases*. **4**(4): e658.

Janssens. P (1996). La trypanosomiase en Angola a` l'aube du 20e sie` cle. Re´ flexions sur les e´ pide´ mies des bassins du Cuanza et du Congo. [Trypanosomiasis in Angola at the dawn of the 20th Century. Reflections on the epidemics of Cuanza and Congo basins.] Bulletin des se´ ances — Acade´ mie royale des sciences d'outre-mer ;**3**:537-69. [In French]

Jenni. L and Brun. R (1982). A new in vitro test for human serum resistance of *Trypanosoma* (T.) *brucei*. *Acta Trop*. **39**: 281–284.

Jenni. L *et al* (1986). Hybrid formation between African trypanosomes during cyclical transmission. *Nature.* **322:** 173–175.

Jiang. Z, Zhang. X, Deka. R and Jin. L (2005). Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acid Research*. **33**(10):1-9.

Jones. C *et al* (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding.* **3**: 381-390.

Jones. D, Taylor. W and Thornton. J (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences.* **8**(3):275-82.

Kabiri. M, Franco. J, Simarro. P, Ruiz. A, Sarsa. M and Steverding. D (1999). Detection of *Trypanosoma brucei gambiense* in sleeping sickness suspects by PCR amplification of expression-site-associated genes 6 and 7. *Tropical Medicine and*

International Health. **4**(10): 658-661.

Kanmogne. G, Bailey. M, Gibson. W (1997). Wide variation in DNA content among *Trypanosoma brucei* ssp. isolates. *Acta Tropica*.**63**:75-87.

Kennedy. P (2006). Human African Trypanosomiasis—Neurological Aspect. *Journal of Neurology*. **253.** 411–416.

Kidwell. M and Lisch. D (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution.* **55**(1):1-24.

Kimmel. B, ole-MoiYoi. K and Young. J (1987). Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Molecular *and Cellular Biology*. **7**(4): 1465-1475.

Koerner. T, De Raadt. P and Maudlin. I (1995). The 1901 Uganda sleeping sickness epidemic revisited: a case of mistaken identity? *Parasitology Today*. **11**(8):303–306

Koffi M, *et al* (2009). Population genetics *of Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *Proceedings of the National Academy of Sciences.***106**:209–214.

Koffi. M, Solano. P, Barnabe. C, de Meeuˆs. T, Bucheton. B, Cuny. G and Jamonneau. V (2007). Genetic characterisation of *Trypanosoma brucei* s.l. using microsatellite typing: New perspectives for the molecular epidemiology of human African trypanosomosis. *Infection, Genetics and Evolution.* **7**:675-684.

Kofler. R, Betancourt. A and Schlötterer. C (2012) Sequencing of Pooled DNA Samples (Pool- Seq) Uncovers Complex Dynamics of Transposable Element Insertions in *Drosophila melanogaster*. *PLOS Genetics.* **8**(1) e1002487-e1002487

Komba. E, Kibona. S, Ambwene. A, Stevens. J and Gibson. W (1997). Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. *Parasitology.* **115**(6): 571–579.

Kouadio. I, Sokouri. D, Koffi. M *et al* (2014). Molecular Characterization and Prevalence of *Trypanosoma* Species in Cattle from a Northern Livestock Area in Côte d'Ivoire. *Open Journal of Veterinary Medicine*. **4**(12): 314-321.

Kuzoe. F (1993). Current Situation of African Trypanosomiasis. *Acta Tropica*. **54**:153-162.

Labrador. M and Corces. V (1997). Transposable element-host interactions: regulation of insertion and excision. *Annual Review of Genetics.* **31;** 381–404.

Langlands. B (1967). The sleeping sickness epidemic of Uganda, 1900–1920: a study in historical geography. Kampala, Makerere University College, (Occasional Paper No. 1).

Larkin. M, Blackshields. G, Brown. N *et al* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**(21): 2947-2948.

Lasken. R and Egholm. M (2003). Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends in Biotechnology*. **21**(12):531-5.

Lehane. M, Masangi. A, Whitaker. C and Lehane. S (2000). Grouping of trypanosome species in mixed infections in *Glossina pallidipes*. *Parasitology*. **120**:583-592.

Lejon. V, Jacobsb. J and Simarro. P (2013). Elimination of sleeping sickness hindered by difficult diagnosis. *Bull World Health Organization*. **91**:718.

Lejon. V, Reiber. H, Legros. D, Dje. N, Magnus. E, Wouters. I, Sindic. C and Büscher. P (2003). Intrathecal Immune Response Pattern for Improved Diagnosis of Central Nervous System Involvement in Trypanosomiasis. *Journal of Infectious Diseases*. **187**:1475–1483.

Levin. N and Delano. S (2011) Evaluation and Treatment of *Malassezia*-Related Skin Disorders. *Cosmetic Dermatology*. **24**(3): 137-145.

Loparev. V, Cartas. M, Monken. C, Velpandi. A. and Srinivasan. A (1991). An efficient and simple method of DNA extraction from whole blood and cell lines to identify infectious agents. *Journal of Virological Methods*. **34**; 105-112.

Louis. F, Simarro P and Les. D (2005). Difficult start for the campaign against sleeping sickness in French equatorial *Africa Médecine Tropicale* [Article in French] **65**:251–257.

Luckins. A (1992). Methods for the diagnosis of trypanosomiasis in livestock. *World animal review* **70**:15-20.

Lymbery .A and Thompson. R (2012). The Molecular epidemiology of parasite infections: tools and applications. *Molecular & Biochemical Parasitology.* **181**: 102-116.

Lynch. M and Conery. J (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*. **290**:1151-1155.

MacGregor. P and Matthews. K (2010). New discoveries in the transmission biology of sleeping sickness parasites: applying the basics. *Journal of Molecular Medicine.* **88**(9):865–871.

Magnus. E, Vervoort. T and Van Meirvenne. N (1978). A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of *T. b. gambiense* trypanosomiasis. *Ann Soc Belg Med Trop.* **58**:169-176.

Masiga. K, Smyth. J, Hayes. P, Bromidge. J and Gibson. C (1992). Sensitive detection of trypanosomes in tsetse flies by DNA amplification. *International Journal of Parasitology* **22**: 909-918.

Mathieu-Daude. F, Bicart-See. A, Bosseno. M, Breniere. S and Tibayrenc. M (1994). Identification of *Trypanosoma brucei gambiense* group I by a specific kinetoplast DNA probe. *The American Journal of Tropical Medicine and Hygiene.* **50**:13-19.

Matovu. E, Kuepfer. I, Boobo. A, Kibona. S and Burri. C (2010). Comparative Detection of Trypanosomal DNA by Loop-Mediated Isothermal Amplification and PCR from Flinders Technology Associates Cards Spotted with Patient Blood. Journal *of* Clinical Microbiology. **48**(6): 2087–2090.

Matsui. S, KIM. J, Greenberg. H, *et al* (1993). Cloning and characterization of human astrovirus immunoreactive epitopes. *Journal of Virology*. **67**(3): 1712-1715.

McClintock. B (1978). Mechanisms that rapidly reorganize the genome. *Stadler Genetics Symposia.* **10:**25–47.

McVean, G *et al* (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56–65.

Mehlitz. D, Zillmann. U, Scott. C and Godfrey. D (1982) Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. III. Characterization of *Trypanozoon* stocks by isoenzymes and sensitivity to human serum*. Tropenmedizin und Parasitologie.* **33**:113–118.

Moloo. S (1993). The distribution of Glossina species in Africa and their natural hosts. *Insect Science and its Application*. **14**:511–527.

Moore. A and Richer. M (2001). Re-emergence of epidemic sleeping sickness in southern Sudan. *Tropical Medicine and International Health*. **6**:342–347.

Moser. D, Cook .A, Ochs. E, Bailey. P, McKane. R and Donelson. E (1989). Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. *Parasitology*. **99**:57-66.

Mseddi. F, Jarboui. M, Sellami. A, Sellami. H and Ayadi. A (2011). A rapid and easy method for the DNA extraction from *Cryptococcus neoformans*. *Biological Procedures Online*. **13**(5):1-3.

Namangala. B *et al* (2012). The use of Loop-mediated Isothermal Amplification (LAMP) to detect the re-emerging Human African Trypanosomiasis (HAT) in the Luangwa and Zambezi valleys. *Parasites & Vectors* **5**: 282.

Ngotho. M, Kagira. J, Gachie. B, Karanja. S, Waema. M, Maranga. D and Maina. N (2015). Loop Mediated Isothermal Amplification for Detection of *Trypanosoma brucei gambiense* in Urine and Saliva Samples in Nonhuman Primate Model. *BioMed Research International*. [Online] [Accessed on 10/03/15] Available; http://www.hindawi.com/journals/bmri/aa/867846/ref/

Njiokou. F, Nimpaye. H, Simo. G, *et al* (2010). Domestic animals as potential reservoir hosts of *Trypanosoma brucei gambiense* in sleeping sickness foci in Cameroon. *Parasite*. **17**(1):61–66.

Njiru. Z, Ndung'u. K, Matete. G, Ndungu. J and Gibson. W (2004). Detection of *Trypanosoma brucei rhodesiense* in animals from sleeping sickness foci in East Africa using the serum resistance associated (*SRA*) gene. *Acta Trpoica* **90**:249- 254.

Njiru. Z, Constantine. C, Ndung'u. J, Robertson. I, Okaye. S, Thompson. R and Reid. S (2004). Detection of *Trypanosoma evansi* in camels using PCR and CATT/*T. evansi* tests in Kenya, *Veterinary Parasitology*. **124** (3–4): 187-199.

Njiru. Z, Mikosza. A, Matovu. E, Enyaru. J, Ouma. J, Kibona. S, Thompson. Rc, Ndung'u. J (2008). African trypanosomiasis: Sensitive and rapid detection of the sub-genus *Trypanozoon* by loop-mediated isothermal amplification (LAMP) of parasite DNA. *International Journal of Parasitology.* **38** (5): 589-599.

Njiru. Z, Constantine. C, Guya. S, Crowther. J, Kiragu. M, Thompson. C and Dávila. M (2005). The use of ITS1 rDNA PCR in detecting pathogenic African trypanosomes. *Parasitology Research*. **95**:186-192.

Njiru. Z, Gitonga. K and Ndungu. K (2011). The typing of *Trypanosoma evansi* isolates using mobile genetic element (MGE) PCR. Parasitology Research. **108**(6):1583-587.

Notomi. T, Okayama. H, Masubuchi. H, Yonekawa. T, Watanabe. K, Amino. N and Hase. T (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acid Research*. **28**(12): e63.

Oberle. M, Balmer. O, Brun. R and Roditi. I (2010). Bottlenecks and the Maintenance of Minor Genotypes during the Life Cycle of *Trypanosoma brucei*. *PLoS Pathog.* **6**(7):e1001023.

Odiit. M, Kansiime. F and Enyaru. J (1997). Duration of symptoms and case fatality of sleeping sickness caused by *Trypanosoma brucei rhodesiense* in Tororo, Uganda. *East African Medical Journal*, **74**(12):792–795.

Orlandi. P and Lampel. K (2000)**.** Extraction-free, filter-based template preparation for rapid and sensitive PCR detection of pathogenic parasitic protozoa. *Journal Clinical of Microbiology*. **38**(6)**:**2271-2277.

Pei. B, Sisu. C, and Frankish. A *et al* (2012). The GENCODE pseudogene resource. *Genome Biology*. **13**:R51.

Penchenier. L, Simo. G, Grébaut. P, Nkinin. S, Laveissière. C and Herder. S (2000). Diagnosis of human trypanosomiasis, due to *Trypanosoma brucei gambiense* in central Africa, by the polymerase chain reaction. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **94**:392-394.

Pépin. A, Mpia. B and Iloasebe. M (2002*). Trypanosoma brucei gambiense* African trypanosomiasis: differences between men and women in severity of disease and response to treatment. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **96**(4): 421–426.

Pèpin. J and Mèda. A (2001). The Epidemiology and Control of Human African Trypanosomiasis. *Advances in Parasitology*, **49**: 71-132.

Picard-Meyer. E, Barrat. J and Cliquet. F (2007). Use of filter paper (FTA®) technology for sampling, recovery and molecular characterisation of rabies viruses**.** *Journal of Virological Methods.* **140:**174-182.

Picozzi. K, Carrington. M and Welburn. S (2008). A multiplex PCR that discriminates between *Trypanosoma brucei brucei* and zoonotic *T. b. rhodesiense. Experimental Parasitology*. **118**:41-469.

Piskurek, O and Jackson. D (2012). Transposable Elements: From DNA Parasites to Architects of Metazoan Evolution. *Genes.* **3**(3): 409–422.

Polski. J, Kimzey. S, Percival. R and Grosso. L (1998) Rapid and effective processing of blood specimens for diagnostic PCR using filter paper and Chelex-100. *Molecular Pathology*, **51**:215-217.

Radwanska. M, Chamekh. M, Vanhamme. L, Claes. F, Magez. S, Magnus. E, de Baetselier. P, Buscher. P and Pays. E (2002b). The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. *The American Journal of Tropical Medicine and Hygiene.* **67**:684-690.

Radwanska. M, Claes. F, Magez. S, Magnus. E, Perez-Morga. D, Pays. E and Buscher. P (2002a). Novel primer sequences for polymerase chain reaction-based detection of *Trypanosoma brucei gambiense*. *The American Journal of Tropical Medicine and Hygiene* **67**: 289-295.

Raper. J, Portela. M, Lugli. E, Frevert. U and Tomlinson. S (2001). Trypanosome lytic factors: novel mediators of human innate immunity. *Current Opinion in Microbiology.* **4**:402–408.

Reithinger. R, Lambson. B, Barker. D and Davies. C (2000). Use of PCR to detect *Leishmania (*Viannia) spp. in dog blood and bone marrow. *Journal of Clinical Microbiology*. **38**: 748-751.

Rickman. L and Robson. J (1970). The testing of proven *Trypanosoma brucei* and *T. rhodesiense* strains by the blood incubation infectivity test. *Bulletin of the World Health Organization* **42**: 911–916.

Robays. J, Bilengue. M, Van der Stuyft. P and Boelaert. M (2004). The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo. *Tropical Medicine & International Health.* **9**: 542-550.

Robays. J, Ebeja Kadima. A, Lutumba. P, *et al* (2004). Human African trypanosomiasis amongst urban residents in Kinshasa: a case-control study. *Trop Med Int Health*.; **9**(8):869–875.

Roy. C and Deo. I (2014). Gene duplication: A major force in evolution and bio-diversity. *International Journal of Biodiversity and Conservation*. **6**(1):41-49.

Roy-Engel. A, Carroll. M, *et al* (2001). *Alu* Insertion Polymorphisms for the Study of Human Genomic Diversity. *Genetics* **159**: 279–290.

Sacramento. D, Badrane. H, Bourhy. H and Tordo. N (1992). Molecular epidemiology of rabies virus in France: comparison with vaccine strains. *Journal of General Virology*. **73**:1149-1158.

Schares. G and Mehlitz. D (1996). Sleeping sickness in Zaire: a nested polymerase chain reaction improves the identification of *Trypanosoma* (*Trypanozoon*) *brucei gambiense* by specific kinetoplast DNA probes. *Tropical Medicine & International Health.* **1**: 59-70.

Schütt. I and Mehlitz. D (1981). On the persistence of human serum resistance and isoenzyme patterns of *Trypanozoon* in experimentally infected pigs. *Acta Tropica*. **38**(4):367–373.

Screaton. G, Bangham. C and Bell. J (1993). Direct sequencing of single primer PCR products: a rapid method to achieve short chromosomal walks *Nucleic Acids Res*. **21**: 2263–2264.

Sharma. R, Gluenz. E, Peacock. L, Gibson. W, Gull. K and Carrington. M (2009). The heart of darkness: growth and form of *Trypanosoma brucei* in the tsetse fly. *Trends Parasitol*ogy .**25**(11):517–524.

Shedlock. A, Takahashi. K and Okada. N (2003). SINEs of speciation: tracking lineages with retroposons. *TRENDS in Ecology and Evolution*.**19:** 545-553.

Simarro.P, Sima. O, Mir. M, Mateo. MJ and Roche. J (1991) La luttecontre la trypanosomiase humaine africaine dans le foyer de Lubaen Guinée Equatoriale: bilan de trois méthodes. *Bulletin del'Organisation Mondiale de la Santé* **69**, 451–457.

Simarro.P, Ruiz. J. A, Franco. J. R and Josenando. T (1999) Attitude towards CATT-positive individuals without parasitological confirmation in the African Trypanosomiasis (*T.b. gambiense*) focus of Quiçama (Angola). *Tropical Medicine & International Health*, **4**: 858–861.

Simarro.P, Diarra. A, Ruiz Postigo. A, Franco. J and Jannin. J (2011). The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000–2009: The Way Forward. *PLOS Neglected Diseases.* **5**(7): 1-7.

Simarro.P, Cecchi, G, Franco. JR, Paone. M, Fèvre. EM, Diarra. A, *et al* (2012) Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis*. **6**(10):e1859

Simarro. P, Franco. J, Diarra. A, Ruiz Postigo. J and Jannin. J (2013). Diversity of human African trypanosomiasis epidemiological settings requires fine-tuning control strategies to facilitate disease elimination. *Research and Reports in Tropical Medicine*. **4**:1–6.

Simo. G, Njiokou. F, Tume. C, Lueong. S and De Meeûs. T (2010). Population genetic structure of Central African *Trypanosoma brucei gambiense* isolates using microsatellite DNA markers. *Infection, Genetics and Evolution*. **10** (1):68-76.

Simo. G, Mbida. J, Eyenga. V, Asonganyi. T, Njiokou. F and Pascal Grébaut. P (2014). Challenges towards the elimination of Human African Trypanosomiasis in the sleeping sickness focus of Campo in southern Cameroon. *Parasites and Vectors*. **7**:374

Singer-Sam. J, Tanguay. R and Riggs. A (1989) Use of Chelex to improve the PCR signal from a small number of cells. *Amplifications: A forum for PCR users*, 11.

Solano. P, Jamonneau. V, N'Guessan. P, N'Dri. L, Dje. N, Miezan. T, Lejon. V, Büscher. P and Garcia. A (2002). Comparison of different DNA preparation protocols for PCR diagnosis of Human African Trypanosomosis in Côte d'Ivoire. *Acta Tropica*. **82**(3):349-56.

Stanghellini. A & Josenando. T (2001). The situation of sleeping sickness in Angola: a calamity. *Tropical Medicine and International Health* **6**, 330–334.

Strøma. G, Marit. G, Hanevik. K *et al* (2014). Comparison of four methods for extracting DNA from dried blood on filter paper for PCR targeting the mitochondrial Plasmodium genome. *Royal Society of Tropical Medicine and Hygiene.* **108**: 488– 494.

Sudarshi. D, Lawrence. S, Pickrell. O, Eligar. V, Walters. R, Quaderi. S *et al* (2014). Human African Trypanosomiasis Presenting at Least 29 Years after Infection—What Can This Teach Us about the Pathogenesis and Control of This Neglected Tropical Disease? *PLoS Neglected Tropical Diseases*, **8**(12), e3349.

Tait. A (1980). Evidence for diploidy and mating in trypanosomes. *Nature*. **287**:536–538.

Tait. A, Barry. J, Wink. R, Sanderson. A and Crowe. J (1985). Enzyme variation in *T. brucei* ssp. II. Evidence for *T. b. rhodesiense* being a set of variants of *T. b. brucei*. Parasitology. **90**:89-100.

TDR. Disease Burden and Epidemiological Trends. Geneva, 2002.

TDR. Strategic Review of Traps and Targets for Tsetse and African Trypanosomiasis Control. Geneva, 2004.

Terry. R, Smith. J, Duncanson. P and Hide. G (2001)**.** MGE-PCR: a novel approach to the analysis of *Toxoplasma gondii* strain differentiation using mobile genetic elements. *International Journal of Parasitology.* **31**: 155-161.

The Arabidopsis Genome Iniativ. (2000). *Nature.* **404:** 711-815.

Thekisoe. O, Kuboki. N, Nambota. A, Fujisaki. K, Sugimoto. C, Igarashi. I, Yasuda. J and Inoue. N (2007). Species-specific loop-mediated isothermal amplification (LAMP) for diagnosis of trypanosomosis. *Acta Tropica*. **102**: 182–189.

Thomas. M, Macias. F, Alonso. C and Lopex. M (2010). The Biology and Evolution of Transposable elements in Parasites. *Cell.* **26**; 350-362.

Thompson. R and Conlan. J (2011). Emerging issues and parasite zoonoses in the SE Asian and Australasian region. *Veterinary Parasitology.* **181**(1): 69–73.

Tilley. A, Hide. G, Fevre. E and Welburn. S (2003). *Trypanosoma brucei*: Trypanosome strain typing using PCR analysis of mobile genetic elements (MGE-PCR).
Experimental Parasitology. **104**:26-32.

Traub. R, Monis. P and Robertson. I (2005). Molecular Epidemiology: a Multisiciplinary Approach to Understanding Parasitic Zoonoses. *International Journal for Parasitology* **35.** 1295-1307.

Truc. P, Jamonneau. V, Cuny. G and Frézil. J (1999). Use of polymerase chain reaction in human African trypanosomiasis stage determination and follow-up. *Bulletin of the World Health Organization*. **77**(9): 745–748.

Truc, p., Grébaut, p., Lando, a., Makiadi Donzoau, f., Penchenier, l., Herder, s.,

Geiger, a., Vatunga, g. and Josenando, t (2011). Epidemiological aspects of the transmission of the parasites causing human African trypanosomiasis in Angola. *Annals of Tropical Medicine & Parasitology.* **105**(3). 261-265.

Truc, P., Tiouchichine, M. L., Cuny, G., Vatunga, G., Josenando, T., Simo, G. and Herder, S (2012). Multiple infections of *Trypanosoma brucei gambiense* in blood and cerebrospinal fluid of human African trypanosomosis patients from Angola: Consequences on clinical course and treatment outcome. *Infection, Genetics and Evolution* **12**. 399–402.

Tu. Z and Li. S (2000). Mobile Genetic Elements of Malaria Vectors and Other Mosquitoes. In: Madame Curie Bioscience Database [Online]. Austin (TX): Landes Bioscience; 2000-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK6186/

United Nations Department of Public Information (1995) The United Nations and the Situation in Angola. UN Reproduction Section, New York.

Uzureau. P *et al* (2013). Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. *Nature.* **501**: 430-434.

Vickerman. K. (1985). Developmental Cycles and Biology of Pathogenic Trypanosomes. *British Medical Bulletin,* **41**(2):105-114.

Vos. P, Hogers, R, Bleeker. M, *et al* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research.* **23**: 4407-4414.

Walsh. P, Metzger. D and Higuchi. R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*. **10**:506-513.

Wastling. S and Welburn. S (2011). Diagnosis of human sleeping sickness: sense and sensitivity. *Trends Parasitology.* **27**(9):394-402.

Weekly epidemiological record (2010). Dracunculiasis eradication – global surveillance summary, 2009. **85**(19):166-76.

Welburn. S, Fevre. E, Coleman. P, Odiit. M and Maudlin. I (2001) Sleeping sickness: a tale of two diseases. *Trends in Parasitology* **17**: 19-24.

Welburn. S, Picozzi. K, Fèvre. EM, Coleman. P, Odiit. M, Carrington. M and Maudlin. I (2001). Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. *The LANCET*. **358**(9298). 2017–2019

Welburn. S and Maudlin. I (2012). Priorities for the Elimination of Sleeping Sickness-Chapter 4. *Advances in Parasitology*. **79**: 299–337

WHO (2015). Cases of sleeping sickness drop to lowest level in 75 years. Geneva. [Online] [Accessed on 25/5/15]. Available from: http://www.who.int/trypanosomiasis_african/cases_drop_to_lowest_since_75_years/en/

WHO. Control and surveillance of human African trypanosomiasis: report of a WHO expert committee 2013 [online] [Accessed on 25/1/15]. Available from. http://apps.who.int/iris/bitstream/10665/95732/1/9789241209847_eng.pdf

WHO. Global Health Observatory, Dracunculiasis (guinea-worm disease). [Online] [Accessed on 10/1/15] available: http://www.who.int/gho/neglected_diseases/dracunculiasis/en/

WHO. Report of a WHO meeting on elimination of African trypanosomiasis (*Trypanosoma brucei gambiense*). Geneva. 2012 [Online] [Accessed on 12/11/14] available:

http://apps.who.int/iris/bitstream/10665/79689/1/WHO_HTM_NTD_IDM_2013.4_eng.pdf

WHO. Report on African Trypanosomes (sleeping sickness). Geneva, World Health Organization, 2001. [Online] [Accessed on 12/11/14] available: http://www.who.int/tdr/publications/documents/aftryp_swg.pdf

WHO. The history of sleeping sickness. Fourth International Course on African Trypanosomoses. Tunis. World Health Organization 2005**.** [Online] [Accessed on 20/01/15] available: http://www.who.int/zoonoses/fourth_afry_course05.pdf

Wickstead. B, Ersfeld. K and Keith. G (2003). Repetitive Elements in Genomes of Parasitic Protozoa *Microbiology and Molecular Biology Reviews*. **67**(3): 360–375.

Witherspoon. D, Marchani. E, Watkins. S, *et al* (2006). Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and *Alu* insertions. *Human Heredity* **62**: 30–46.

Woo. P (1970). The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis. *Acta Trop* **27**: 384-386.

Wooden. J, Kyes. S and Sibley. C (1993). PCR and strain identification in *Plasmodium falciparum*. *Parasitology today.* **9**(8): 303-305.

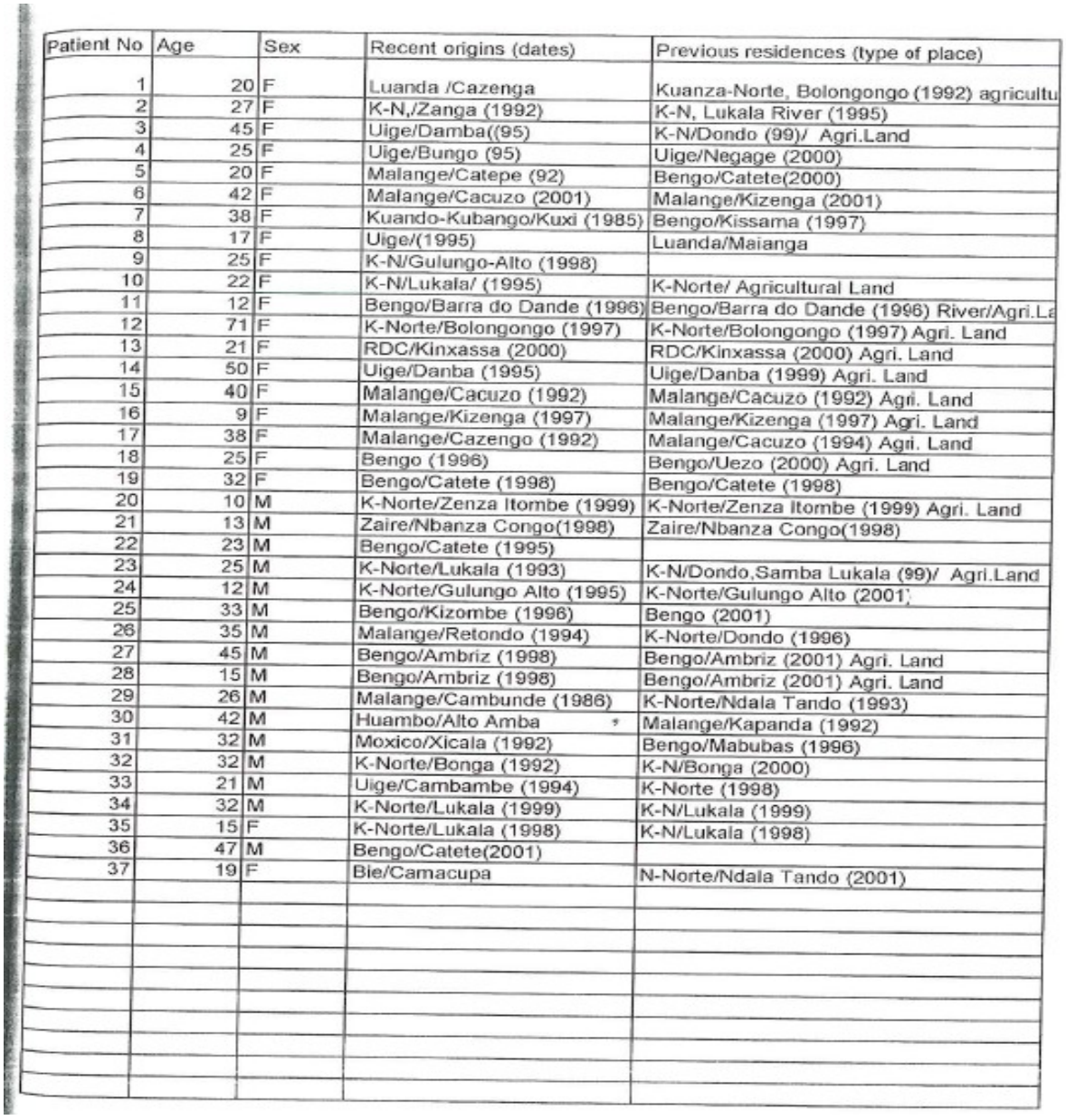
World Bank (2014) Country profile Angola. [online] http://www.worldbank.org/en/country/angola

Yang. Z, Nielsen. R, Goldman. N and Pedersen. A (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* **155**(1):431-49

Appendices

Appendix 1

Figures below illustrate the full information obtained from the source of the collection



Patient No	Age	Sex	Recent origins (dates)	Previous residences (type of place)
1	20	F	Luanda /Cazenga	Kuanza-Norte, Bolongongo (1992) agricultu
2	27	F	K-N,/Zanga (1992)	K-N, Lukala River (1995)
3	45	F	Uige/Damba((95)	K-N/Dondo (99)/ Agri.Land
4	25	F	Uige/Bungo (95)	Uige/Negage (2000)
5	20	F	Malange/Catepe (92)	Bengo/Catete(2000)
6	42	F	Malange/Cacuzo (2001)	Malange/Kizenga (2001)
7	38	F	Kuando-Kubango/Kuxi (1985)	Bengo/Kissama (1997)
8	17	F	Uige/(1995)	Luanda/Maianga
9	25	F	K-N/Gulungo-Alto (1998)	
10	22	F	K-N/Lukala/ (1995)	K-Norte/ Agricultural Land
11	12	F	Bengo/Barra do Dande (1996)	Bengo/Barra do Dande (1996) River/Agri.La
12	71	F	K-Norte/Bolongongo (1997)	K-Norte/Bolongongo (1997) Agri. Land
13	21	F	RDC/Kinxassa (2000)	RDC/Kinxassa (2000) Agri. Land
14	50	F	Uige/Danba (1995)	Uige/Danba (1999) Agri. Land
15	40	F	Malange/Cacuzo (1992)	Malange/Cacuzo (1992) Agri. Land
16	9	F	Malange/Kizenga (1997)	Malange/Kizenga (1997) Agri. Land
17	38	F	Malange/Cazengo (1992)	Malange/Cacuzo (1994) Agri. Land
18	25	F	Bengo (1996)	Bengo/Uezo (2000) Agri. Land
19	32	F	Bengo/Catete (1998)	Bengo/Catete (1998)
20	10	M	K-Norte/Zenza Itombe (1999)	K-Norte/Zenza Itombe (1999) Agri. Land
21	13	M	Zaire/Nbanza Congo(1998)	Zaire/Nbanza Congo(1998)
22	23	M	Bengo/Catete (1995)	
23	25	M	K-Norte/Lukala (1993)	K-N/Dondo,Samba Lukala (99)/ Agri.Land
24	12	M	K-Norte/Gulungo Alto (1995)	K-Norte/Gulungo Alto (2001)
25	33	M	Bengo/Kizombe (1996)	Bengo (2001)
26	35	M	Malange/Retondo (1994)	K-Norte/Dondo (1996)
27	45	M	Bengo/Ambriz (1998)	Bengo/Ambriz (2001) Agri. Land
28	15	M	Bengo/Ambriz (1998)	Bengo/Ambriz (2001) Agri. Land
29	26	M	Malange/Cambunde (1986)	K-Norte/Ndala Tando (1993)
30	42	M	Huambo/Alto Amba	Malange/Kapanda (1992)
31	32	M	Moxico/Xicala (1992)	Bengo/Mabubas (1996)
32	32	M	K-Norte/Bonga (1992)	K-N/Bonga (2000)
33	21	M	Uige/Cambambe (1994)	K-Norte (1998)
34	32	M	K-Norte/Lukala (1999)	K-N/Lukala (1999)
35	15	F	K-Norte/Lukala (1998)	K-N/Lukala (1998)
36	47	M	Bengo/Catete(2001)	
37	19	F	Bie/Camacupa	N-Norte/Ndala Tando (2001)

Figure (1). The collected samples from patients on FTA card from Angola. Illustrates the Age, Gender, previous and the sample collection locations

No family history
Treated (98) melarsoprol, NFH
NFH
Treated (4/4/01) 30 days melarsoprol, NFH
NFH, abdomal body moviment, released from H ospital
treated (01/2001) melarsoprol, NFH
SS present in family
Treated (1995) melarsoprol
NFH
SS present in family
Treated (09/01) melarsoprol, NFH
Treated (1997) melarsoprol(30 days) , NFH
NFH
NFH
SS in family
NFH
NFH
SS present in family
Treated (2000), NFH
NFH
NFH
NFH
NFH, Treated (99), (2001)
NFH
NFH
NFH, has been in different provinces of Angola while in the army
SS in family
nefew of the above
NFH, has been in different provinces of Angola while in the army
NFH, has been in different provinces of Angola while in the army
NFH, has been in different provinces of Angola while in the army
NFH, has been treated (19/2/2001) melarsoprol (10days)
NFH, has been treated (4/4/2000) melarsoprol (10days)
NFH, has been in a coma for 2 weeks before samples were teken
NFH
NFH, has tested negative for trypanossome
NFH, has tested negative for trypanossome

Figure **(2).** The medical history of the participant patients in this study

Cattle and tsetse	Ocupation	reasons for hospitalization
Present	Farmer	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present		Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Not Present	student	Trypanosomiasis
Present	house wife	Trypanosomiasis
Not Present	student	Trypanosomiasis
Present	student	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Not Present		Trypanosomiasis
Not Present	house wife	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Not Present	student	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	house wife	Trypanosomiasis
Present	student	Trypanosomiasis
Present	student	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	soldier	Trypanosomiasis
Present		Trypanosomiasis
Present	teacher	Trypanosomiasis
Present	soldier	Trypanosomiasis
Present	Farmer	Trypanosomiasis
Present	student	Trypanosomiasis
Present	soldier	Trypanosomiasis
Present	soldier	Trypanosomiasis, tiphoid fe
Present	soldier	Trypanosomiasis
Present	Farmer	Trypanosomiasis/relapse
Present	traveller/se	Trypanosomiasis/relapse
Present		suspected tryp
Present	student	not hospitalised
Present	Farmer	not hospitalised
Present	student	not hospitalised

Figure **(3).** The occupation of the participant patients in this study.

Pre-Treat(3days)	Tretment	Diagnosis	
none	none	CATT, LNP, LP, TIT	
Yes	Pentamidine (10 days), saline intravenouselly(3d)	CATT, LNP, LP,TIT	
Yes	none@ time	CATT, LNP, LP	
Yes	melarsoprol (8 days)	CATT, LNP, LP	
Yes	melarsoprol (10 days)	CATT, LNP, LP	
Yes	melasoprol (10 days)	CATT, LNP	
Yes	melarsoprol (10 days)	CATT, LNP	
none	none	CATT, LNP, LP, TIT	
Yes	none@time	CATT, LNP, LP, TIT	
none	none@time	CATT, LNP, LP, TIT	
Yes	melarsoprol (1 day @ time)	CATT, LNP, LP, TIT	
Yes	melarsoprol (6days@ time)	CATT, LNP, LP, TIT	
Yes	none@ time	CATT, LNP, TIT	
Yes	melarsoprol (10 days)	CATT, LNP, TIT	
Yes	melarsoprol (10 days)	CATT, LNP, TIT	
Yes	melarsoprol (10 days)	CATT, LNP, TIT	
Yes	none@ time	CATT, LNP, TIT	
Yes	melarsoprol (10 days) D.F.M.O	CATT, LNP, LP, TIT	
Yes	melarsoprol (10 days)	CATT, LNP, LP, TIT	
Yes	melarsoprol (10 days)	CATT, LNP, LP, TIT	
Yes	melarsoprol (7 days)	CATT, LNP, LP, TIT	
Yes	melarsoprol (2 days @ time)	CATT, LNP, LP, TIT	
Yes	none@ time	CATT, LNP, TIT	
Yes	melarsoprol (10days), Saline and complex B vitamin	CATT, LNP, TIT	
Yes	melarsoprol (10days), Saline and D.F.M.O	CATT, LNP, LP, TIT	
Yes	melarsoprol (2 days @ time)	CATT, LNP, TIT	
Yes	Pentamidine (10 days)	CATT, LNP, TIT	
Yes	melarsoprol (4day@ time)	CATT, LNP, TIT	
Yes	melarsoprol (4day@ time)	CATT, LNP, TIT	
Yes	melarsoprol (2day@ time)	CATT, LNP, TIT	
Yes	none@ time	CATT, LNP, TIT	
none	none@ time	CATT, LNP, TIT, LCR	
none	none@ time	CATT, LNP, TIT, LCR,LP	
none	none@ time	CATT, LNP, TIT, LCR	
none	none@ time	CATT, LNP, TIT,	
none	none@ time	CATT, LNP, TIT,	
none	none@ time	CATT, LNP, TIT,	
lymph node puncture			
lymph node puncture			
lymph node puncture			

Figure **(4).** The treatment information of the patients and the initial diagnosis at the arrival at the hospital.

Symptoms and signs	Time of symptoms
headace, dizzness, stomach pain	2 Y
headace, dizzness, stomach pain, lost weigth	3 Y
headace, dizzness, stomach pain, lost weigth sleeliping	2Y
diharrea, sleeping,dizziness,lost weigth, paralysis of body, uncompreensive tolk	3y
headace,dizziness, mental disturbance, fever	2y
headace,dizziness, diharrea, fever	1y
headace,apetite lost,sleeping, trembeling, speech loss,auditionloss	3Y
headace,apetite lost,lost weigth, fainting	7Y
sleeping, headace,mental disturbance,trembeling	3Y
lost speech, sleeping, headace	1Y
sleeping, headace,mental disturbance,trembeling	2Y
sleeping, headace,mental disturbance,trembeling	4Y
headace, loss weigth, tiredness	2Y
headace, loss weigth, visual disturbances, trembeling	2Y
tiredness, headace, loss weigth, backpain	1Y
sleeping, headace, trembeling	2Y
headace, trembeling, uncontrolled hands moviments	1Y
headace, sleeping	2Y
headace, sleeping, trembeling, skin problem, loss weigth, speech and adition loss	2y
sleeping, skin problems, weigth loss, headace, trembeling	2Y
cramps, headace, sleeping, skin problem, diharreia	2Y
sleeping, headace, weigth loss, trembeling	1Y
headace, cramps, diharrea	3Y
speech loss, sleeping,dehidrated, apetite loss	1Y
headace, tiredness, trembeling	1Y
headace, tiredness, weigth loss	1Y
sleeping, headace,	2Y
headace	1Y
headace, fever, apetite loss	1Y
headace, fever, sleeping, apetite loss, paralysis of legs	2Y
headace, stomach ace, apetite loss, paralysis of legs	6 mounths
mental disturbances, headace	2Y
lost speech, sleeping, headace	1Y
paralysis of body, speech loss, weigth loss	1Y
tiredness, headace, skin lesions	6mounths
headace,vomiting, fever, annemia	2 Y
headace	3mounths

Figure **(5).** The symptoms and the duration of the symptoms of the participant patients.

Appendix 2

The multiple alignment of the amino acid sequences of the RHS/pseudogenes present in TREU927 genome. Illustrating two features of the RHS genes subfamilies

```
RHS1                KGVEYVFKSDVYIRREALRVWYKVENVLNRWLMDEVIVDDASNVLIGTPGIGKSFSVGSL
Tb927.1.180         KGVEYVFKSDVYIRREALRVWYKVENVLNRWLMDEVIVDDASNVLIGTPGIGKSFSVGSL
Tb927.1.220         KGVEYVFKSDVYIRREALRVWYKVENVLNRWLMDEVIVDDASNVLIGTPGIGKSFSVGSL
Tb927.2.100         KGVGYVFNSDVYIRRETLRVWHKVKERLNQWLMGEVAINPMFHVVIGTPGIGKSFSVGSL
Tb927.4.110#*       KGVGYVFNSDVYIRRETLRVWHKVEERLNQWLMGEVAINPMFHVVIGTPGIGKSFSVGSL
Tb927.2.370         KGVEYVFKSDVYIRREALRVWYKVENVLNRWLMDEVIVDDASNVLIGTPGIGKSFSVGSL
Tb927.4.200         KGVEYVFKSDVYIRRETLRVWHKVEERLNKWLMGEVTVNPMFHVLIGTPGIGKSFSTAPF
Tb927.6.160         KGVEYVFNDDVYIRRETLRVWHKVEERLNKWLMGEVIVNPMAHVLIGTPGIGKSFSTAPF
Tb927.4.100#        ------------------------------------------------------- I

RHS2                 RVWYLVKHRLDLWLGPGVNKKQIPCVLVGTPGIGKSFGVGSFLLQRLLHYDSELLRNIAY
Tb927.2.1160#*       ------------------------------------------------------------
Tb927.4.220#*        ------------------------------------------------------MNQQVPI
Tb927.2.400          RVWYLVKHRLDLWLGPGVNKKQIPCVLVGTPGIGKSFGVGSFLLQRLLHYDSELLRNIAY
Tb927.2.280          RVWYLVKHRLDLWLGPGVNKKQIPCVLVGTPGIGKSFGVGSFLLQRLLHYDSELLRNIAY
Tb927.4.210#*        ------------------------------------NPAPRVAALRFGAVADYRL
Tb927.10.110#*       ------------------------MAAAVGGNSHEGTKKKKILLQRLLHYDSELLRIIAC
Tb927.2.380          RVWYLVKQRLDLWLGPELVEEPFPCVLVGTPGIGKSFGVGSFLLQRLLHYDSERLRNIAY
Tb927.2.1180         RNDAEPPVVQQRVENNNQPQWGLFSCVDAVLLNGLPHPRNMMLNDFLRRNFGRRYNVNEE
Tb927.6.5160         RNDAEPPAVQQRVENNKQPQWTILSSVEAVLLNGLPHPGNMSVNDFLRFNLLEDIDLDQR
Tb927.4.280          PPPAQPPQMPQRYKRFS--DWKMNSTVRDFLLEEYATLPKMNLHDFLNQCFSNTYNT---
Tb927.4.290          PPPAQPPQMPQRYKRFS--DWKMNSTVRDFLLEEYATLPKMNLHDFLNQCFSNTYNT---
Tb927.4.270          PPPAQPPQMPQRYKRFS--DWKMNSTVRDFLLERYATLPKMNLHDFLNQCFSNTYNT---
Tb927.4.260          PPPAQPPQMPQRYKRFS--DCKMNSTVRDFLLEEYATLPKMNLHDFLNQCFSNTYNT---
Tb927.4.300          PPPAQPPQMPQRYKRFS--DWKMNSTVRDFLLEEYATLPKMNLHDFLNQCFSNTYNT---
Tb927.5.280^*        -------------------WVLGSNVKDVLLEEYAGLKDMKLHEFLNQHIGETLGTP--
Tb927.6.150          -------------------------------NETALFESVFLKHVQLGECSSGCVCNPG
Tb927.2.1150#*       ----------------FCSKGCCITIRSCCGISLILRARHIFSSGRQTIIREGLYSMKR

RHS3                 REIVRVWYKVEKDLKTWWVEKTAHRPPIHIVIGTPGIGKSYGLGSFLLHSLLHFHEGMLD
Tb927.2.340          REIMRVWYIIQRGLQTWWLNDTAHRPPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGMLD
Tb927.2.470          REIMRVWYIIQRGLQTWWVEKTARRTPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGMLD
Tb927.2.450          REIMRVWYIIQRGLQAWWVEKTAQRTPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGMLD
Tb927.1.120          REIMRVWYIIQRGLQGRWVEK-AEITPRHIVIGTPGIGKSYGVGSFLLHSLLHFNDGMLD
Tb927.2.1210         REIMRVWYIIQRGLQGRWVEKADITP-RHIVIGTPGIGKSYGVGSFLLHSLLHFHEGMFD
Tb927.2.560          REIMRVWYIIQRGLQGRWVEKAEITP-RHIVIGTPGIGKSYGVGSFLLHSLLHFNDGMLD
Tb927.1.70           REIMRVWYKVEKGLRACWVEKAE-DSPIHIVIGTPGIGKSCGLGSFLLHSLLHFHEGMLD
Tb927.4.240          REIMRVWYIIQRGLQGRWVEKAQ-RPPIHIVIGTPGIGKTCGLGSFLLHSLLHFNEGMLD
Tb927.2.510          REIMRVWYIIQRGLQTWWVERTARITPRHIVIGTPGIGKSYGVGSFLLYSLLHFHDGMFN


RHS4                 IRREIMRVWYIIQRGLQTWWLNDTAHRPPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.2.340          IRREIMRVWYIIQRGLQTWWLNDTAHRPPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.6.5170^*       IRREIMRVWYIIQKDLKAWWVEKTAHRPPIHIVIGTPGIGKSCGVGSFLLH---------
Tb927.2.450          IRREIMRVWYIIQRGLQAWWVEKTAQRTPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.2.470          IRREIMRVWYIIQRGLQTWWVEKTARRTPIHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.1.120          IRREIMRVWYIIQRGLQGRWVEKAEITP-RHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.2.560          IRREIMRVWYIIQRGLQGRWVEKAEITP-RHIVIGTPGIGKSYGVGSFLLHSLLHFNDGM
Tb927.2.1210         IRREIMRVWYIIQRGLQGRWVEKADITP-RHIVIGTPGIGKSYGVGSFLLHSLLHFHEGM
Tb927.6.5180^#*#     ------------------------------------------MLHSLLHFHEGMM
Tb927.1.70           IRREIMRVWYKVEKGLRACWVEKAE-DSPIHIVIGTPGIGKSCGLGSFLLHSLLHFHEGM
Tb927.4.240          IRREIMRVWYIIQRGLQGRWVEKAQ-RPPIHIVIGTPGIGKTCGLGSFLLHSLLHFNEGM
Tb927.2.510          IRREIMRVWYIIQRGLQTWWVERTARITPRHIVIGTPGIGKSYGVGSFLLYSLLHFHDGM

RHS5                 WNVVRNGIELWRTQRVIPGCLYLPRSYVAIGNPGIGKSQNLGSFILYKLLHYDAEELPV
Tb927.1.420          WNVVRNGIELWRTQRVIPGCLYLPRSYVAIGNPGIGKSQNLGSFILYKLLHYDAEELPV
Tb927.2.1170         WNFVRSGIELWRAQRVIPGCPYLPRSYVAIGNPGIGKSQNLGSFILYKLLHYDAEELPV
Tb927.4.250          WNVVRSGIELWRTQRVIPGCLYLPRSYVAIGNPGIGKSQNLGSFILYKLLHYDAEELPV
Tb927.6.140          WNVVRNGIELWRTQRVIPGCPYLPRSYVAIGNPEIGKSQNSGSFILYKLLHYDAEDLPI
Tb927.2.240          WNVVRNGIELWRTQRVMPRCLYFPRSYLAIGNPGIGKSQNLGSFILYKLLHYDAEELPI
Tb927.2.1080         WNVVRSGIELWRTQRVMPRCLYFPRSYLAIGNPGIGKSQNFGSFILYKLLHYDAEELPV
```

```
RHS6                RTQRALPGCLHFPRAYLAIGNPGIGKSQNLGSFILYNLLHYDAEELPVVAYFRGVAAYIF
Tb927.11.4770*^     RTQRALPGCLYFPRSYVAIGNPGIGKSQNLGSFILYNLLHYDAEELPVVAYFRGVAAYIF
```

Figure **(6)**. The multiple alignment of amino acid sequences of *RHS* genes present in TREU927 and *RHS* genes families (1-6). The alignments was carried according to the each *RHS* gene family. The GTP motif is highlighted in red. The *RHS* genes that possess a mobile elements are marked with (^) for ingi insertion, (#) RIME B and (*) RIME A. The alignment was carried in EBI clustalw.

```
Tb927.2.280           TPGIGKSFGVGSFLLQRLLHYDSELLRNIAYFVKGRAYIFFRATNDHPGKVVFYEKEKDG  360
Tb927.2.400           TPGIGKSFGVGSFLLQRLLHYDSELLRNIAYFVKGRAYIFFRATNDHPGKVVFYEKEKDG  360
Tb927.2.380           TPGIGKSFGVGSFLLQRLLHYDSERLRNIAYFVKGEAYIFFRATNDHPRKVVFYEKEVDG  360
Tb927.10.110#*        GGNSHEGTKKKKILLQRLLHYDSELLRIIACFVKGGAYIFFRATNGHPGRVVFYE-ENDG  64
Tb927.4.210#*         ------------NPAPRVAALRFGAVADYRLFCEGGIYFLHGDKRSSGKGCILKRKWSGC  48
Tb927.1.180           TPGIGKSFSVGSLLLYKLLHYEASQLQIIIYVVRGKAYVFHKPIGGRAGYVTFYNDYGNA  398
Tb927.1.220           TPGIGKSFSVGSLLLYKLLHYEASQLQIIIYVVRGKAYVFHKPIGGRAGYVTFYNDYGNA  398
Tb927.2.370           TPGIGKSFSVGSLLLYKLLHYEASQLQIIIYVVRGKAYVFHKPIGGRAGYVTFYNDYGNA  399
Tb927.2.100           TPGIGKSFSVGSLLLYKLLHYEASQLQIIIYVVRGKAYVFHKPIGGRAGYVTFYSNYENA  394
Tb927.4.100#          ------------ILLYKLLHYEASQLQIIIYVVEGEAYVFRKPKGDRAGYVTFYSNYKSA  48
Tb927.4.200           TPGIGKSFSTAPFLLYKLLHYESSQLQIIIYVVRGKAYVFHKPKGGRAGYVTFYSNYEQA  243
Tb927.6.160           TPGIGKSFSTAPFLLYKLLHYEASQLQIIIYVVRGKAYVFHKPIGGRAGYVTFYSNYEHA  243
Tb927.4.110#*         TPGIGKSFSVGSLLLYT-------------------------------------------  359
Tb927.1.70            TPGIGKSCGLGSFLLHSLLHFHEGMLDVVAYFVGKIAYLIYNKKPGEEGRVEEDRSTDSA  235
Tb927.4.240           TPGIGKTCGLGSFLLHSLLHFNEGMLDVVAYFVGKIAYLIHNKKPGEEGRVEEDRSTDSA  359
Tb927.2.510           TPGIGKSYGVGSFLLYSLLHFHDGMFNVIAYFVGKIAYLIYNKKPGEEGRVERYKVPEDA  393
Tb927.2.1210          TPGIGKSYGVGSFLLHSLLHFHEGMFDVVAYFTDGIAYLIYNKKPGEEGRVEEYEDVKAA  384
Tb927.6.5180^#*#      -------------MLHSLLHFHEGMLDVVAYFTGGIAYLIYNDKPGEEGRVEQYKGVTAS  47
Tb927.2.560           TPGIGKSYGVGSFLLHSLLHFNDGMLDVVAYFVGKIAYLIYNKKPGEGGKVEQYEDVKAA  388
Tb927.6.5160          TAGIGKSCGVGSFLLHSLLHFHEGMLDVVAYFTDTIAFLIYNRKGDERGRVERYEGSRAA  382
Tb927.6.5170^*        TPGIGKSCGVGSFLLH--------------------------------------------  337
Tb927.2.1180          TPGIGKSYGLGSFLLHSLLHFHEGMLDVVAYFTDTIAYLIYNRKGDERGRVVRYEYLRAA  381
Tb927.4.220#*         TAGIGKSVGVGSFLLQ--------------------------------------------  343
Tb927.2.340           TPGIGKSYGVGSFLLHSLLHFNDGMLDVVAYFTDGNAFLIYNKKPGEEERVVQYEHVKAA  381
Tb927.2.470           TPGIGKSYGVGSFLLHSLLHFNDGMLDVVAYFVGKIAYLIYNKKPGEEGRVERYKVPEDA  382
Tb927.2.450           TPGIGKSYGVGSFLLHSLLHFNDGMLDVVAYFVGEIAFLIYNKKPGEEGRVEKYVVSKNA  381
Tb927.1.120           TPGIGKSYGVGSFLLHSLLHFNDGMLDVVAYFTDGIAYLIYNKKPGEEGRVERYKFPEDA  381
Tb927.4.280           TSGIGKTCGLGSFLFYSLLHFNEGMLDVVAYFLEDAAFLIYNREGEGRGTVDEYKRWESA  359
Tb927.4.290           TSGIGKTCGLGSFLFYSLLHFNEGMLDVVAYFLEDAAFLIYNREGEGRGTVDEYKRWESA  359
Tb927.4.270           TSGIGKTCGLGSFLFYSLLHFNEGMLDVVAYFLEDAAFLIYNREGEGRGTVDEYKRWESA  359
Tb927.4.260           TSGIGKTCGLGSFLFYSLLHFNEGMLDVVAYFLEDSAFLIYNREGEERGTVDEYERWESA  359
Tb927.4.300           TSGIGKTCGLGSFLFYSLLHFNEGMLDVVAYFLEDAAFLIYNREGEERGTVDEYERWESA  374
Tb927.5.280^*         TPGIGKYGLGLFFTSF--------------------------------------------  276
Tb927.2.1170          NPGIGKSQNLGSFILYKLLHYDAEELPVVAYFRGVAAYIFEKSGGGGVGRVREYS--KEA  233
Tb927.4.250           NPGIGKSQNLGSFILYKLLHYDAEELPVVAYFRGVAAYIFEKSDNGGVGRVREYS--KEA  233
Tb927.1.420           NPGIGKSQNLGSFILYKLLHYDAEELPVVAYFRGVAAYIFEKSSGGGVGRVKKYS--KGA  233
Tb927.6.140           NPEIGKSQNSGSFILYKLLHYDAEDLPIAAYFCGEVAYIFEKLGEGGVGIIRQYG--KGA  170
Tb927.2.240           NPGIGKSQNLGSFILYKLLHYDAEELPIVAYFCGVAAYIFEKSGGGGVGRVREYS--KEA  233
Tb927.2.1080          NPGIGKSQNFGSFILYKLLHYDAEELPVVAYFRGVAAYIFEKSGGGGVGRVREYS--KEA  233
Tb927.11.4770*^       NPGIGKSQNLGSFILYNLLHYDAEELPVVAYFRGVAAYIFEKSGGGGVGRVREYS--KEA  182
Tb927.2.1160#*        SAGIGKSVGAGY------------------------------------------------  327
Tb927.6.150           SGGFTDTYCYWYTWYWIMWPWIIFASFIASLPRNACCCVFHR-------RDSLLNIQKEK  271
Tb927.2.1150#*        IESVGGHSMYARQLPSEGFSWER-----------------RAGQIKCGHICLHDGKCCA  267
```

Figure (7). The amino acid sequences multiple alignment of *RHS* genes present in TREU927. The insertion site sequence for mobile element is underlined, the *RHS* genes that possess a mobile elements are marked with (^) for ingi insertion, (#) RIME B and (*) RIME A.

Appendix 3

```
Tb927.4.290       ------------------------------------------------------------
Tb927.4.280       ------------------------------------------------------------
Tb927.4.270       ------------------------------------------------------------
Tb927.4.260       ------------------------------------------------------------
Tb927.4.300       ATGTCAAGAGCCAATAGCCCCGCAGCTCCACAAGGAAATAACGCAAACCCACAAGTGCCA 60
Tb927.2.1180      ----------------------------------------ATGAATCAACAGGTTCCT 18


Tb927.4.290       ------------------AT------GAGGAGGCCACGGGATGAACATGTAGCACCTCCT 36
Tb927.4.280       ------------------AT------GAGGAGGCCACGGGATGAACATGTAGCACCTCCT 36
Tb927.4.270       ------------------AT------GAGGAGGCCACGGGATGAACATGTAGCACCTCCT 36
Tb927.4.260       ------------------AT------GAGGAGGCCACGGGATGAACATGTAGCACCTCCT 36
Tb927.4.300       AATAACATTGAAGGGCCCAT------GAGGAGGCCACGGGATGAAAATGTAGCACCTCCT 114
Tb927.2.1180      ATTGAAGGAAGGGGTGACATTGAGGGAAGGAGACGAGAAAATGAGGAGGCTGCGCGTAAC 78
                            **       ***** * *    ****  * *  ** * *


Tb927.4.290       CCTGCTCAGCCTCCACAGATGCCGCAGAGA-TATAAAAGATTTTCTGATTGGAAAATGAA 95
Tb927.4.280       CCTGCTCAGCCTCCACAGATGCCGCAGAGA-TATAAAAGATTTTCTGATTGGAAAATGAA 95
Tb927.4.270       CCTGCTCAGCCTCCACAGATGCCGCAGAGA-TATAAAAGATTTTCTGATTGGAAAATGAA 95
Tb927.4.260       CCTGCTCAGCCTCCACAGATGCCGCAGAGA-TATAAAAGATTTTCTGATTGTAAAATGAA 95
Tb927.4.300       CCTGCTCAGCCTCCACAGATGCCGCAGAGA-TATAAAAGATTTTCTGATTGGAAAATGAA 173
Tb927.2.1180      GATGCGGAACCTCCAGTAGTGCAGCAACGAGTGGAGAATAACAACCAACC--ACAATGGG 136
                    ***  * ******    *** ***  ** *  * ** *    *  *    * ****


Tb927.4.290       -------CAG-TACAGTGAGGGATTTTTTGCTGGAAGAATATGCGACATTGCCCAAA-AT 146
Tb927.4.280       -------CAG-TACAGTGAGGGATTTTTTGCTGGAAGAATATGCGACATTGCCCAAA-AT 146
Tb927.4.270       -------CAG-TACAGTGAGGGATTTTTTGCTGGAAAGATATGCGACATTGCCCAAA-AT 146
Tb927.4.260       -------CAG-TACAGTGAGGGATTTTTTGCTGGAAGAATATGCGACATTGCCCAAA-AT 146
Tb927.4.300       -------CAG-TACAGTGAGGGATTTTTTGCTGGAAGAATATGCGACATTGCCCAAA-AT 224
Tb927.2.1180      GTCTTTTCAGCTGCGTTGACGCAGTTTT-GCTG-AATGGTCTGCCACATCCTAGAAATAT 194
                         *** * *  *** * * **** **** **   * *** ****     *** **


Tb927.4.290       GAATCTGCATGACTTTTTGAATCAATGCTTCT--CAAACACGTATAACAC-AGTGAA---  200
Tb927.4.280       GAATCTGCATGACTTTTTGAATCAATGCTTCT--CAAACACGTATAACAC-AGTGAA---  200
Tb927.4.270       GAATCTGCATGACTTTTTGAATCAATGCTTCT--CAAACACGTATAACAC-AGTGAA---  200
Tb927.4.260       GAATCTGCATGACTTTTTGAATCAATGCTTCT--CAAACACGTATAACAC-AGTGAA---  200
Tb927.4.300       GAATCTGCATGACTTTTTGAATCAATGCTTCT--CAAACACGTATAACAC-AGTGAA---  278
Tb927.2.1180      GATGTTGAATGATTTTTTGAGGCG-TAATTTTGGCAGA-AGATATAATGTGAATGAAGAA  252
                  **   ** **** *******  *  *  ** *  ** * *  *****    * ****


Tb927.4.290       -----TGTGTCAATGGAAACGTTTGTTAAGAATCCTGAAGATTATATTAAGGATGCCGAG  255
Tb927.4.280       -----TGTGTCAATGGAAACGTTTGTTAAGAATCCTGAAGATTATATTAAGGATGCCGAG  255
Tb927.4.270       -----TGTGTCAATGGAAACGTTTGTTAAGAATCCTGAAGATTATATTAAGGATGCCGAG  255
Tb927.4.260       -----TGTGTCAATGGAAACGTTTGTTAAGAATCCTGAAGATTATATTAAGGATGCCGAG  255
Tb927.4.300       -----TGTGTCAATGGAAACGTTTGTTAAGAATCCTGAAGATTATATTAAGGATGCCGAG  333
Tb927.2.1180      AATAATGTATCAATGTCTGCGTTCGTACTGGAACCTGAAGAGTATATTAATGACGTTAAT  312
                       *** ******    **** **   * * ******** ******** ** *   *


Tb927.4.290       ATTCTTGA--GGATATACAGGGGCTTTATGAGTTCAGGGAGTACAAAAGGATTGTTGAGA  313
Tb927.4.280       ATTCTTGA--GGATATACAGGGGCTTTATGAGTTCAGGGAGTACAAAAGGATTGTTGAGA  313
Tb927.4.270       ATTCTTGA--GGATATACAGGGGCTTTATGAGTTCAGGGAGTACAAAAGGATTGTTGAGA  313
Tb927.4.260       ATTCTTGA--GGATATACAGGGGCTTTATGAGTTCAGGGAGTACAAAAGGATTGTTGAGA  313
Tb927.4.300       ATTCTTGA--GGATATACAGGATACGGATGAGTTT--------CAA-----TTACTGAGG  378
Tb927.2.1180      GCGCTCAATCGAATATTC-GCGACAACA-GAGTATAAAGTGTACAAACGTTTTGTCAG--  368
                     **  *  * **** * *       * ****          ***     **


Tb927.4.290       AACTGCCTGAAACAGAAATGTGGATAAACTACCTTAAAGAGGAAAAAAT-TTTTACCATC  372
Tb927.4.280       AACTGCCTGAAACAGAAATGTGGATAAACTACCTTAAAGAGGAAAAAAT-TTTTACCATC  372
Tb927.4.270       AACTGCCTGAAACAGAAATGTGGATAAACTACCTAAAAGAGGAAAAAAT-TTTTACCATC  372
Tb927.4.260       AACTGCCTGAAACAGAAATGTGGATAAACTACCTTAAAGAGGAAAAAAT-TTATACCATC  372
Tb927.4.300       G-----------CAGATATTT-----ATCTCTCT----GAGGAAAAAAT-TTTTACCATC  417
Tb927.2.1180      ------------CGTCTATG-------GTTTCTTTGAGGATGAGGGAATATTGAATCTTC  409
                             *    **          *   *    ** **   *** **  * * **


Tb927.4.290       AGCGATTGGGTGAACGGTGCAACTCCAGAGGTGAA----AAACTCCAT-----TAGCCCC  423
Tb927.4.280       AGCGATTGGGTGAACGGTGCAACTCCAGAGGTGAA----AAACTCCAT-----TAGCCCC  423
Tb927.4.270       AGCGATTGGGTGAACGGTGCAACTCCAGAGGTGAA----AAACTCCAT-----TAGCCCC  423
Tb927.4.260       AGCGATTGGGTGAACGGTGCAACTCCAGAGGTGAA----AAACTCCAT-----TAGCCCT  423
Tb927.4.300       AGCGATTGGGTGAACGGTGCAACTCCAGAAGTGAA----AAACTCCAT-----TAGCCCC  468
Tb927.2.1180      AGCGCT--GGC-AACAA-GCAGATGAGGAAGCAAAGGTCAGGTTACAGGCGGATATTCGT  465
                  **** *  **  ***   ***  *   ** *  **    *   * **      **  *
```

```
Tb927.4.290       GTAGCAAGGG--GAAGG------TTGAAT-GCAGT-AC-AACATGCTG----------TC  462
Tb927.4.280       GTAGCAAGGG--GAAGG------TTGAAT-GCAGT-AC-AACATGCTG----------TC  462
Tb927.4.270       GTAGCAAGGG--GAAGG------TTGAAT-GCAGT-AC-AACATGCTG----------TC  462
Tb927.4.260       GTAGCAAGGG--GAAGG------TTGAAT-GCAGT-AC-AATATGCTG----------TC  462
Tb927.4.300       GTAGCAAGGG--GAAGG------TTGAAT-GCAGT-AC-AACATGCTG----------TC  507
Tb927.2.1180      GGACTGAGAGATGGAGAGCGCCTTTGGATTGTTGTCACTAACATGTTGAATGATGCACTC  525
                  * *   ** *  * **       *** ** *  ** ** ** *** **          **

Tb927.4.290       ACGGAATCAGAAGTATG-GTTTCT---GCCGCGTGT--TTCTGTCAAAAGTTGTCGAGAT  516
Tb927.4.280       ACGGAATCAGAAGTATG-GTTTCT---GCCGCGTGT--TTCTGTCAAAAGTTGTCGAGAT  516
Tb927.4.270       ACGGAATCAGAAGTATG-GTTTCT---GCCGCGTGT--TTCTGTCAAAAGTTGTCGAGAT  516
Tb927.4.260       ACGGAATCAGAAGTATG-GTTTCT---GCCGCGTGT--TTCTGTCAAAAGTTGTCGAGAT  516
Tb927.4.300       ACGGAATCAGAAGTATG-GTTTCT---GCCGCGTGT--TTCTGTCAAAAGTTGTCGAGAT  561
Tb927.2.1180      AATGAAGCAAGGGAACGAGCTGCTCAAACAGCTTGTAGTGCTGTCGAA--CTGAAAGGAT  583
                  *  *** **   * * * * * **    * ** ***  * ***** **   **    ***

Tb927.4.290       GTTTACGATTCTTTTTACAATGCGAAATGGAGTTATGTGATGTCGGGTTATGATGCGGAG  576
Tb927.4.280       GTTTACGATTCTTTTTACAATGCGAAATGGAGTTATGTGATGTCGGGTTATGATGCGGAG  576
Tb927.4.270       GTTTACGATTCTTTTTACAATGCGAAATGGAGTTATGTGATGTCGGGTTATGATGCGGAG  576
Tb927.4.260       GTTTACGATTCTTTTTACAATGCGAAATGGAGTTATGTGATGTCGGGTTATGATGCGGAG  576
Tb927.4.300       GTTTACGATTCTTTTTACAATGCGAAATGGAGTTATGTGATGTCGGGTTATGATGCGGAG  621
Tb927.2.1180      -TGTATGAATCCATATATAATGCGAAATGGAGTTATGTGATGTCGGGTTATCATACAGAG  642
                   * ** ** **  * ** ********************************** ** * ***


Tb927.4.290       CCACTTGGAATGAAGGTATTCGATGGAAGGCCGCAGCGCATATGGACGGAGGCAGAATTG  636
Tb927.4.280       CCACTTGGAATGAAGGTATTCGATGGAAGGCCGCAGCGCATATGGACGGAGGCAGAATTG  636
Tb927.4.270       CCACTTGGAATGAAGGTATTCGATGGAAGGCCGCAGCGCATATGGACGGAGGCAGAATTG  636
Tb927.4.260       CCACTTGGAATGAAGGTATTCGATGGAAGGCCGCAGCGCATATGGACGGAGGCAGAATTG  636
Tb927.4.300       CCACTTGGAATGAAGGTATTCGATGGAAGGCCGCAGCGCATATGGACGGAGGCAGAATTG  681
Tb927.2.1180      CCACTTGGCATGAAAGTGTTCGATGGAAGGCCGCAGCGCATGTGGACCGAAGAAGATGTG  702
                  ******** ***** ** ********************** ***** ** * ***  **

Tb927.4.290       AATATAACTCCTAAAACTGCCGATGTTGATGCAAGAGTAAAGGAAGGATCCAATGGTTTG  696
Tb927.4.280       AATATAACTCCTAAAACTGCCGATTTTGATGCAAGAGTAAAGGAAGGATCCAATGGTTTG  696
Tb927.4.270       AATATAACTCCTAAAACTGCCGATGTTGATGCAAGAGTAAAGGAAGGATCCAATGGTTTG  696
Tb927.4.260       AATATAACTCCTAAAACTGCCGATGTTGATGCAAGAGTAAAGGAAGGATCCAATGGTTTG  696
Tb927.4.300       AATATAACTCCTAAAACTGCCGATTTTGATGCAAGAGTAAAGGAAGGATCCAATGGTTTG  741
Tb927.2.1180      GATATAACTCCTCTTCCTGCGAATGTTGATGCAGAGATTGAGGAGAGACCCGATGGTTTG  762
                   ***********    ****  ** ********    *  ****  ** ** ********

Tb927.4.290       GAGATTTTTGTTCTTACGTCGGAGAAGGGTTGGCCATACAACAGGTTTGTATTGGGTTAC  756
Tb927.4.280       GAGATTTTTGTTCTTACGTCGGAGAAGGGTTGGCCATACAACAGGTTTGTATTGGGTTAC  756
Tb927.4.270       GAGATTTTTGTTCTTACGTCGGAGAAGGGTTGGCCATACAACAGGTTTGTATTGGGTTAC  756
Tb927.4.260       GAGATTTTTGTTCTTACGTCGGAGAAGGGTTGGCCATACAACAGGTTTGTATTGGGTTAC  756
Tb927.4.300       GAGATTTTTGTTCTTACGTCGGAGAAGGGTTGGCCATACAACAGGTTTGTATTGGGTTAC  801
Tb927.2.1180      GAGATTTTTGTTCTCACTTCGGAGAAAGGTTGGCCATACAACAGGTTTGCATTGGATTAC  822
                  ************** ** ******** ********************** ***** ****

Tb927.4.290       ACTGAGAGATGCAAGACAATTTGCAAGCATGTATATATCCGTCGTGAAATTATGCGCGTG  816
Tb927.4.280       ACTGAGAGATGCAAGACAATTTGCAAGCATGTATATATCCGTCGTGAAATTATGCGCGTG  816
Tb927.4.270       ACTGAGAGATGCAAGACAATTTGCAAGCATGTATATATCCGTCGTGAAATTATGCGCGTG  816
Tb927.4.260       ACTGAGAGATGCAAGACAATTTGCAAGCATGTATATATCCGTCGTGAAATTATGCGCGTG  816
Tb927.4.300       ACTGAGAGATGCAAGACAATTTGCAAGCATGTATATATCCGTCGTGAAATTATGCGCGTG  861
Tb927.2.1180      ACTACGGGGCGCAAAGTAGTATTTCAGCACGTATACATCCGTCGTGAAATTGTGCGCGTG  882
                  ***  * *  ****   * * *   **** ***** *************** ********

Tb927.4.290       TGGTATATAATCCAACGAGGTCTGAAAGCATGGTGGATGGAAGCAACAGCTGAGGGTCCA  876
Tb927.4.280       TGGTATATAATCCAACGAGGTCTGAAAGCATGGTGGATGGAAGCAACAGCTGAGGGTCCA  876
Tb927.4.270       TGGTATATAATCCAACGAGGTCTGAAAGCATGGTGGATGGAAGCAACAGCTGAGGGTCCA  876
Tb927.4.260       TGGTATATAATCCAACGAGGTCTGAAAGCATGGTGGATGGAAGCAACAGCTGAGGGTCCA  876
Tb927.4.300       TGGTATATAATCCAACGAGGTCTGAAAGCATGGTGGATGGAAGCAACAGCTGAGGGTCCA  921
Tb927.2.1180      TGGTATAAAGTCGAAAAAGATCTGAAAACATGGTGGGTGGAAAAGACAGCTCACAGACCA  942
                  ******* * ** **  ** ******* ******** *****   ****** *  * ***

Tb927.4.290       CCAATACATATTATTATTGGCACGTCTGGAATTGGTAAAACATGTGGCCTTGGATCATTT  936
Tb927.4.280       CCAATACATATTATTATTGGCACGTCTGGAATTGGTAAAACATGTGGCCTTGGATCATTT  936
Tb927.4.270       CCAATACATATTATTATTGGCACGTCTGGAATTGGTAAAACATGTGGCCTTGGATCATTT  936
Tb927.4.260       CCAATACATATTATTATTGGCACGTCTGGAATTGGTAAAACATGTGGCCTTGGATCATTT  936
Tb927.4.300       CCAATACATATTATTATTGGCACGTCTGGAATTGGTAAAACATGTGGCCTTGGATCATTT  981
Tb927.2.1180      CCGATACATATTGTTATTGGTACACCTGGTATTGGTAAATCATATGGCCTTGGATCATTT  1002
                  ** ********* ******* **  **** ********* *** ****************

Tb927.4.290       TTATTTTATTCACTGCTTCACTTCAATGAAGGAATGCTTGATGTTGTTGCATATTTCTTG  996
Tb927.4.280       TTATTTTATTCACTGCTTCACTTCAATGAAGGAATGCTTGATGTTGTTGCATATTTCTTG  996
Tb927.4.270       TTATTTTATTCACTGCTTCACTTCAATGAAGGAATGCTTGATGTTGTTGCATATTTCTTG  996
Tb927.4.260       TTATTTTATTCACTGCTTCACTTCAATGAAGGAATGCTTGATGTTGTTGCATATTTCTTG  996
```

```
Tb927.4.300       TTATTTTATTCACTGCTTCACTTCAATGAAGGAATGCTTGATGTTGTTGCATATTTCTTG  1041
Tb927.2.1180      TTGCTTCATTCGTTGCTTCACTTCCATGAAGGAATGCTTGATGTTGTTGCATATTTCACA  1062
                  **  ** ****  ********** ********************************

Tb927.4.290       GAGGATGCTGCCTTCCTGATATACAACCGAGAGGGTGAGGGAAGAGGGACTGTTGATGAA  1056
Tb927.4.280       GAGGATGCTGCCTTCCTGATATACAACCGAGAGGGTGAGGGAAGAGGGACTGTTGATGAA  1056
Tb927.4.270       GAGGATGCTGCCTTCCTGATATACAACCGAGAGGGTGAGGGAAGAGGGACTGTTGATGAA  1056
Tb927.4.260       GAGGATTCTGCCTTCCTGATATACAACCGAGAGGGTGAGGAAAGAGGGACTGTTGATGAA  1056
Tb927.4.300       GAGGATGCTGCCTTCCTGATATACAACCGAGAGGGTGAGGAAAGAGGGACTGTTGATGAA  1101
Tb927.2.1180      GACACGATCGCCTACTTAATATATAACAGAAAGGGTGATGAAAGAGGGAGGGTTGTGCGG  1122
                  **       **** * * ***** *** ** ******* * ********  ****

Tb927.4.290       TACAAACGCTGGGAGTCTGCCGTTGCTGCCATAAATAAAATGAAATCTGAAAGCAAAGGG  1116
Tb927.4.280       TACAAACGCTGGGAGTCTGCCGTTGCTGCCATAAATAAAATGAAATCTGAAAGCAAAGGG  1116
Tb927.4.270       TACAAACGCTGGGAGTCTGCCGTTTCTGCCATAAATAAAATGAAATCTGAAAGCAAAGGG  1116
Tb927.4.260       TACGAACGCTGGGAGTCTGCCGTTGCTGCCATAAATAAAATGAAATCTGAAAGCAAAGGG  1116
Tb927.4.300       TACGAACGCTGGGAGTCTGCCGTTGCTGCCATAAATAAAATGAAATTTGAAAACAGAGGA  1161
Tb927.2.1180      TACGAATATTTGAGAGCTGCGGTCAACGCCATAAATAAAATGAAATTTGAAAACAGAGGA  1182
                  *** **   * *    **** **    ****************** ***** ** ***

Tb927.4.290       CACATTGTTATGGACACAAGGTATGCAATGGAGGAACTATACACTCAACTTCCCTCTGAT  1176
Tb927.4.280       CACATTGTTATGGACACAAGGTATGCAATGGAGGAACTATACACTCAACTTCCCTCTGAT  1176
Tb927.4.270       CACATTGTTATGGACACAAGGTATGCAATGGAGGAACTATACACTCAACTTCCCTCTGAT  1176
Tb927.4.260       CACATTGTTATGGACACAAGGTATGCAATGGAGCAACTATACACTCAACTTCCCTCTGAT  1176
Tb927.4.300       CACATTATTATGGACACAAGGTATGCAATGGAGAAACTATACACTCAACTTCCCTCTGAT  1221
Tb927.2.1180      CACATTATTATGGACATAAGGTATGCAATGCAGCAGCTTTACACTCAACTTCCCTCTGAT  1242
                  ****** ********* ************* ** * ** *********************

Tb927.4.290       GTTTGGAGTGTAACTTTCCTCACTTTCCCTAAAAGTGCCCATTTTGATGAGTGGAGTGCA  1236
Tb927.4.280       GTTTGGAGTGTAACTTTCCTCACTTTCCCTAAAAGTGCCCATTTTGATGAGTGGAGTGCA  1236
Tb927.4.270       GTTTGGAGTGTAACTTTCCTCACTTTCCCTAAAAGTGCCCATTTTGATGAGTGGAGTGCA  1236
Tb927.4.260       GTTTGGAGTGTAACTTTCCTCACTTTCCCTAAAAGTGACCATTTTGATGAGTGGAGTGCA  1236
Tb927.4.300       GTTTGGAGTGTAACTTTCCTCACTTTCCCTAAAAGTGCCCATTTTGATGAGTGGAGTGCA  1281
Tb927.2.1180      GTTTGGAGTGTAACTATTTTCACTTCCCCGAATAGTTCCCACTTTGGTAACTGGACTACA  1302
                  *************** *  ****** *** ** ***  *** **** * * **** * **

Tb927.4.290       AATAAGGGAGGCCACCATATTATTATAAACTGCGATGATGTGCGTGACATGAAGGCATTT  1296
Tb927.4.280       AATAAGGGAGGCCACCATATTATTATAAACTGCGATGATGTGCGTGACATGAAGGCATTT  1296
Tb927.4.270       AATAAGGGAGGCCACCATATTATTATAAACTGCGATGATGTGCGTGACATGAAGGCATTT  1296
Tb927.4.260       AATAAGGGAGGCCACCAAATTATTATGAACTGCGATGATGTGCGTGACATGAAGGCATTT  1296
Tb927.4.300       AATAAGGGAGGCCACCAAATTATTATGAACTGCGATGATGTGCGTGACATGAAGGCATTT  1341
Tb927.2.1180      ATTACGGGAGGCCGCCAAATTATTATAAACTGTGATGATGTGCGTGACATGAAGGCATTT  1362
                  * ** ******** *** ******** ***** ***************************

Tb927.4.290       ATGGCATGGAAAAAGCTGTCAATACATACAGGAGAAAAGGTGTCCAACAGAAGAAGATAC  1356
Tb927.4.280       ATGGCATGGAAAAAGCTGTCAATACATACAGGAGAAAAGGTGTCCAACAGAAGAAGATAC  1356
Tb927.4.270       ATGGCATGGAAAAAGCTGTCAATACATACAGGAGAAAAGGTGTCCAACAGAAGAAGATAC  1356
Tb927.4.260       ATGGCATGGAAAAAGCTGTCAATACATACAGGAGAAAAGGTGTCCAACAGAAGAAGATAC  1356
Tb927.4.300       ATGGCATGGAAAAAGCTGTCAATACATACAGGAGAAAAGGTGTCCAACAGAAGAAGATAC  1401
Tb927.2.1180      GTGGCATGGAAAAAACTGTCAATACATACAAGGGAAAAGGTATCCAACAGAAGAAGACAC  1422
                   ************* *************** * ******** *************** **

Tb927.4.290       GAATTGAAGAGAGAAATTGAAGATGAGTGGAAAATAGTGGAGGGACGCATTAATTTAATT  1416
Tb927.4.280       GAATTGAAGAGAGAAATTGAAGATGAGTGGAAAATAGTGGAGGGACGCATTAATTTAATT  1416
Tb927.4.270       GAATTGAAGAGAGAAATTGAAGATGAGTGGAAAATAGTGGAGGGACGCATTAATTTAATT  1416
Tb927.4.260       GAATTGAAGAGAGAAATTGAAGATGAGTGGAAAATAGTGGAGGAACGCATTAATTTAATT  1416
Tb927.4.300       GAATTGAAGAGAGAAATTGAAGATGAGTGGAAAATAGTGGAGGGACGCATTAATTTAATT  1461
Tb927.2.1180      GAATTGAGGAAGGAAATGGAAGATGAGTGGAGAATAGTGGAGGGACGTATTAATTCAATT  1482
                  ******* **  ***** ************* ************ *** ******* ****

Tb927.4.290       GGACCGCTGCCTCGTTATATATTTGGTTTAGGTTGTTATGAGTGGCATCTGAAAAGTGTG  1476
Tb927.4.280       GGACCGCTGCCTCGTTATATATTTGGTTTAGGTTGTTATGAGTGGCATCTGAAAAGTGTG  1476
Tb927.4.270       GGACCGCTGCCTCGTTATATATTTGGTTTAGGTTGTTATGAGTGGCATCTGAAAAGTGTG  1476
Tb927.4.260       GGACCGCTGCCTCGTTATATATTTGGTTTAGGTTGTTATGAGTGGCATCTGAAAAGTGTG  1476
Tb927.4.300       GGACCGCTGCCTCGTTATATATTTGGTTTAGGTTGTTATGAGTGGCATCTGAAAAGAGTG  1521
Tb927.2.1180      GGGCTTTTGCCTCGTTATATATTTGTTTTAAGCTGTTATGAATGGCGTCTGAAAAGAGTT  1542
                  ** *   ****************** **** * ******** **** ********* **

Tb927.4.290       TATGATGCGTTGAGAGCTATGAAGAAATCTGATAAATATTGCTATAATGATATTATTGAA  1536
Tb927.4.280       TATGATGCGTTGAGAGCTATGAAGAAATCTGATAAATATTGCTATAATGATATTATTGAA  1536
Tb927.4.270       TATGATGCGTTGAGAGCTATGAAGAAATCTGATAAATATTGCTATAATGATATTATTGAA  1536
Tb927.4.260       TATGATGCGTTGAGAGCTATGAAGAAATCTGATAAATATTGCTATAATGATATTATTGAA  1536
Tb927.4.300       TATGATGCGTTGAGAGCTATGAAGAAATCTGATAAATATTGCTATAATGATATTATTGAA  1581
Tb927.2.1180      CATGATGCGTTGGAAACTATGAAGAAGTCGGATGAATATTCCTATAATGATATTATTGAG  1602
                   ***********  * ********** ** *** ****** ******************
```

```
Tb927.4.290       CACACGGCTGACTGGCGGAACGATGAAGTTATGAACAAGTTGGTGAAGATTGTAAGAGTA  1596
Tb927.4.280       CACACGGCTGACTGGCGGAACGATGAAGTTATGAACAAGTTGGTGAAGATTGTAAGAGTA  1596
Tb927.4.270       CACACGGCTGACTGGCGGAACGATGAAGTTATGAACAAGTTGGTGAAGATTGTAAGAGTA  1596
Tb927.4.260       CACACGGCTGACTGGCGGAACGATGAAGTTATGAACAAGTTGGTGAAGATTGTAAGAGTA  1596
Tb927.4.300       CACACGGCTGACTGGCGGAACGATGAAGTTATGAACAAGTTGGTGAAGATTGTAAGAGTA  1641
Tb927.2.1180      CATACCGCTGCCTGGAAGAACAATGAAGTTACGGAAAAGTTGGTAAAGGTTGTAAGAGTA  1662
                  ** ** **** ****  **** ********* * * ******** *** ***********

Tb927.4.290       AAAGGAAATGTTGGAGGAGTCGAATCATACAAATGTCAAGCCCTATCACTGATGATCAGA  1656
Tb927.4.280       AAAGGAAATGTTGGAGGAGTCGAATCATACAAATGTCAAGCCCTATCGCTTATGATCAGA  1656
Tb927.4.270       AAAGGAAATGTTGGAGGAGTCGAATCATACAAATGTCAAGCCCTATCGCTGATGATCAGA  1656
Tb927.4.260       AAAGGAAATGTTGGAGGAGTCGAATCATACAAATGTCAAGCCCTATCGCTGATGATCAGA  1656
Tb927.4.300       AAAGGAAATGTTGGAGGAGTCGAATCATACAAATGTCAAGCCCTATCGCTGATGATCAGA  1701
Tb927.2.1180      AAGGGAAATGTTGGATTTATCGAATCATTTAAATTTCAAGCTCTATCACTGATGATTCGA  1722
                  ** ************    *********  **** ****** ***** ** *****  **

Tb927.4.290       AATATGATGATGAATTAA 1674
Tb927.4.280       AATATGATGATGAATTAA 1674
Tb927.4.270       AATATGATGATGAATTAA 1674
Tb927.4.260       AATATGATGATGAATTAA 1674
Tb927.4.300       AATATGATGATGAATTAA 1719
Tb927.2.1180      AATATGATGATGAGTTAA 1740
                  ************ ****
```

Figure **(8).** Nucleotide sequence alignments the duplicated genes, of RHS/pseudogenes present in TREU927 genome. The adopted mutations at the N-terminal, is underlined.